U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

PB-280 410

# Growth Effects of Major Land Use Projects (Wastewater Facilities). Volume I. Model Specification and Causal Analysis

Abcor, Inc., Wilmington, Mass.

Prepared for

Environmental Protection Agency, Research Triangle Park, N.C.

Mar 78

PB 280 410

EPA-450/3-78-014a

March 1978

# GROWTH EFFECTS OF MAJOR LAND USE PROJECTS (WASTEWATER FACILITIES)

## Volume I: Model Specification and Causal Analysis

U.S. ENVIRONMENTAL PROTECTION AGENCY
Office of Air and Waste Management
Office of Air Quality Planning and Standards
Research Triangle Park, North Carolina 27711

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U. S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

## TECHNICAL REPORT DATA

*(Please read Instructions on the reverse before completing)*

| | | |
|---|---|---|
| 1. REPORT NO.<br>EPA-450/3-78-014a | 2. | 3. RECIPIENT'S ACCESSION NO.<br>PB280410 |
| 4. TITLE AND SUBTITLE<br>Growth Effects of Major Land Use Projects, Wastewater Facilities, Volume I, Model Specification and Causal Analysis | | 5. REPORT DATE<br>March 1978 |
| | | 6. PERFORMING ORGANIZATION CODE |
| 7. AUTHOR(S)<br>Peter H. Guldberg, Ralph B. D'Agostino<br>Richard D. Cunningham | | 8. PERFORMING ORGANIZATION REPORT NO.<br>C-921 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Walden Division of Abcor, Inc.<br>850 Main Street<br>Wilmington, MA 01887 | | 10. PROGRAM ELEMENT NO. |
| | | 11. CONTRACT/GRANT NO.<br>68-02-2594 |
| 12. SPONSORING AGENCY NAME AND ADDRESS<br>Environmental Protection Agency<br>Office of Air Quality Planning and Standards<br>Strategies and Air Standards Division MD-12<br>Research Triangle Park, NC 27711 | | 13. TYPE OF REPORT AND PERIOD COVERED<br>Final |
| | | 14. SPONSORING AGENCY CODE<br>200/04 |

15. SUPPLEMENTARY NOTES

16. ABSTRACT

Growth Effects of Major Land Use Projects is a research program whose goal is to develop methodologies to predict the total air pollution emissions resulting from the construction and operation of major land use projects. Emissions are quantified from the major project, from land use induced by the major project, from secondary activity occurring off-site (e.g., electrical generating stations), and from motor vehicle traffic associated with both the major project and its induced land uses.

This report documents the development of a causal model for the induced land use from wastewater major projects. The report discusses the theoretical basis of the model, the specification of an initial causal model for growth effects, sample selection and data collection, and the testing and refinement of the causal model. A subsequent report will document predictive equations and worksheets for applying the model.

17. KEY WORDS AND DOCUMENT ANALYSIS

| a. DESCRIPTORS | b. IDENTIFIERS/OPEN ENDED TERMS | c. COSATI Field/Group |
|---|---|---|
| Land Use<br>Planning<br>Sewage Treatment Plants | Path Analysis<br>Causal Analysis<br>Secondary Effects<br>Induced Land Use | |

| | | |
|---|---|---|
| 18. DISTRIBUTION STATEMENT<br>Unlimited | 19. SECURITY CLASS *(This Report)*<br>Unclassified | 21. NO. OF PAGES<br>255 |
| | 20. SECURITY CLASS *(This page)*<br>Unclassified | 22. PRICE<br>A12-A01 |

EPA Form 2220-1 (9-73)

EPA-450/3-78-014a

# GROWTH EFFECTS OF MAJOR LAND USE PROJECTS (WASTEWATER FACILITIES)

## Volume I: Model Specification and Causal Analysis

by

Peter H. Guldberg, Ralph B. D'Agostino, and Richard D. Cunningham

Walden Division of Abcor, Inc.
850 Main Street
Wilmington, MA 01887

Contract No. 68-02-2594

EPA Project Officer: Thomas McCurdy

Prepared for

U.S. ENVIRONMENTAL PROTECTION AGENCY
Office of Air and Waste Management
Office of Air Quality Planning and Standards
Research Triangle Park, North Carolina 27711

March 1978

ia

This report is issued by the Environmental Protection Agency to report technical data of interest to a limited number of readers. Copies are available free of charge to Federal employees, current contractors and grantees, and nonprofit organizations - as supplies permit - from the Land Use Planning Office, Office of Air Quality Planning and Standards, Environmental Protection Agency, Research Triangle Park, North Carolina 27711; or, for a nominal fee, from the National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22161.

This report was furnished to the Environmental Protection Agency by Walden Division of Abcor, Inc., Wilmington, Massachusetts 01887, in fulfillment of Contract No. 68-02-2594. This report has been reviewed by the Land Use Planning Office, EPA and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

Publication No. EPA-450/3-78-014a

## ACKNOWLEDGEMENTS

Special appreciation goes to Mr. Thomas McCurdy, the U.S. Environmental Protection Agency project officer for this study, whose extensive assistance and advice was indispensable in the conduct of the study. Appreciation is also extended to all EPA technical committee members for guidance given throughout the study. The technical committee included representatives of the Office of Transportation and Land Use Policy, the Municipal Construction Division (Facilities Requirements Branch) and the Control Program Development Division.

In addition, we wish to thank the more than one hundred individuals in city/county/regional planning agencies and transportation departments nationwide who cooperated with us during the data collection task and helped provide the data on which this study is based. Their time and cooperation was invaluable.

Urban Systems Research & Engineering, Inc. (USRE) of Cambridge, MA were subcontractors to Walden on this study, assisting in the definition of basic model concepts, infrastructure relationships and exogenous variables, and in the testing and refinement of the causal and predictive models. The cooperation of Dr. James F. Hudson, who guided the USRE technical effort, is deeply appreciated.

Finally, recognition is due the Walden staff who performed the long and difficult task of field data collection - Mr. Mahesh Shah, Ms. Diane Gilbert, Mr. Michael Geraghty and Mr. Keith Kennedy.

# TABLE OF CONTENTS

Section Title Page

I INTRODUCTION . . . . . . . . . . . . . . . . . . . . . . . . 1-1
- A. Study Overview . . . . . . . . . . . . . . . . . . . . . . 1-1
- B. Technical Background . . . . . . . . . . . . . . . . . . . 1-3
  - 1. Federal Wastewater Funding Programs . . . . . . . . 1-3
  - 2. Wastewater Facilities . . . . . . . . . . . . . . . 1-4
  - 3. Land Use Impacts . . . . . . . . . . . . . . . . . 1-8
- C. Problem Statement . . . . . . . . . . . . . . . . . . . . 1-13
- D. General Approach . . . . . . . . . . . . . . . . . . . . 1-14
- E. Glossary of Terms . . . . . . . . . . . . . . . . . . . . 1-17

II PHASE I - INITIAL MODEL SPECIFICATION . . . . . . . . . . . 2-1
- A. Perform Literature Searches . . . . . . . . . . . . . . . 2-1
- B. Aerial Photographic Interpretation . . . . . . . . . . . 2-2
- C. Define Modeling Approach . . . . . . . . . . . . . . . . 2-2
  - 1. Theory of Induced Development . . . . . . . . . . . 2-2
  - 2. Approach to Theory Testing . . . . . . . . . . . . 2-3
  - 3. Model Form and Characteristics . . . . . . . . . . 2-5
  - 4. Technique Selection for Path Analysis . . . . . . . 2-6
- D. Define Induced Development . . . . . . . . . . . . . . . 2-11
- E. Define Major Project . . . . . . . . . . . . . . . . . . 2-17
- F. Define Area of Analysis . . . . . . . . . . . . . . . . . 2-23
- G. Initial Choice of Variables . . . . . . . . . . . . . . . 2-29
- H. Specify Initial Path Analytic Land Use Model . . . . . 2-44
  - 1. RES Equation . . . . . . . . . . . . . . . . . . . 2-58
  - 2. COMM Equation . . . . . . . . . . . . . . . . . . . 2-60
  - 3. OFFICE Equation . . . . . . . . . . . . . . . . . . 2-61
  - 4. MANF Equation . . . . . . . . . . . . . . . . . . . 2-62
  - 5. HIWAYS Equation . . . . . . . . . . . . . . . . . . 2-63
  - 6. EDUC Equation . . . . . . . . . . . . . . . . . . . 2-63
  - 7. REC Equation . . . . . . . . . . . . . . . . . . . 2-64
  - 8. WHOLE Equation . . . . . . . . . . . . . . . . . . 2-65
  - 9. OTHER Equation . . . . . . . . . . . . . . . . . . 2-65

III PHASE II - DATA COLLECTION . . . . . . . . . . . . . . . . 3-1
- A. Case Study Selection . . . . . . . . . . . . . . . . . . 3-1
  - 1. Selection Criteria . . . . . . . . . . . . . . . . 3-1
  - 2. Information Sources . . . . . . . . . . . . . . . . 3-3
  - 3. Final Selection . . . . . . . . . . . . . . . . . . 3-9
- B. Data Collection . . . . . . . . . . . . . . . . . . . . . 3-10
  - 1. Transportation Data . . . . . . . . . . . . . . . . 3-10
  - 2. Survey Forms . . . . . . . . . . . . . . . . . . . 3-18
  - 3. Field Surveys . . . . . . . . . . . . . . . . . . . 3-29
  - 4. Create Computer Data File . . . . . . . . . . . . . 3-35

## TABLE OF CONTENTS (CONTINUED)

| Section | Title | Page |
|---|---|---|
| IV | PHASE III - CAUSAL ANALYSIS | 4-1 |
| | A. Path Analysis | 4-1 |
| | 1. Preselection of Exogenous Variables | 4-2 |
| | 2. Identification | 4-3 |
| | 3. Multicollinearity | 4-3 |
| | 4. Suppressor Variables | 4-7 |
| | 5. Theory Trimming | 4-8 |
| | 6. Comparison of Model Forms | 4-24 |
| | B. Coefficient Stability Analysis | 4-27 |
| | 1. Extreme Value Index | 4-30 |
| | 2. Coefficient of Variation Index | 4-34 |
| | C. Determine Net Causal Effects | 4-37 |
| | 1. Approach | 4-37 |
| | 2. Results and Conclusions | 4-38 |
| | D. VMT Model Validation | 4-51 |
| | 1. Develop Test Criteria | 4-51 |
| | 2. Model Validation | 4-56 |
| V | SUMMARY OF RESULTS AND CONCLUSIONS | 5-1 |
| VI | REFERENCES | 6-1 |
| | APPENDIX A - Path Analysis | A-1 |
| | APPENDIX B - Causal Analysis in GEMLUP Projects | B-1 |
| | APPENDIX C - Aerial Photographic Interpretation | C-1 |
| | APPENDIX D - Bibliography | D-1 |
| | APPENDIX E - Case Study Data* | E-1 |
| | APPENDIX F - Condescriptive Output* | F-1 |
| | APPENDIX G - Simple Correlation Output* | G-1 |
| | APPENDIX H - Statistical Output of the Final Path Analysis* | H-1 |
| | APPENDIX I - One Way Frequency Distributions of Model Coefficient Values from the Jackknifing Analysis* | I-1 |

---

*Appendices E through I are not included in this volume. They may be ordered from the Strategies and Air Standards Division, U.S. Environmental Protection Agency MD-12, Research Triangle Park, NC 27711.

## LIST OF TABLES

| Table | Title | Page |
|---|---|---|
| 2-1 | Initial List of Endogenous Variables | 2-13 |
| 2-2 | Point Estimates of the Construction Period of Interceptor Sewer Projects | 2-21 |
| 2-3 | Construction Grant Information on a Sample of Current Wastewater Projects in Federal Regions II, IV, and VI | 2-24 |
| 2-4 | Definition of Endogenous Model Variables | 2-30 |
| 2-5 | Definition of Exogenous Model Variables | 2-33 |
| 2-6 | Independent Variables for the RES Model Equation | 2-49 |
| 2-7 | Independent Variables for the COMM Model Equation | 2-50 |
| 2-8 | Independent Variables for the OFFICE Model Equation | 2-51 |
| 2-9 | Independent Variables for the MANF Model Equation | 2-52 |
| 2-10 | Independent Variables for the HIWAYS Model Equation | 2-53 |
| 2-11 | Independent Variables for the EDUC Model Equation | 2-54 |
| 2-12 | Independent Variables for the REC Model Equation | 2-55 |
| 2-13 | Independent Variables for the WHOLE Model Equation | 2-56 |
| 2-14 | Independent Variables for the OTHER Model Equation | 2-57 |
| 3-1 | Summary of Survey Questionnaire Returns | 3-5 |
| 3-2 | Sample Survey Questionnaire | 3-6 |
| 3-3 | Case Study Major Projects | 3-7 |
| 3-4 | Selected Variables Found Significant In Urban Transportation Planning | 3-14 |
| 3-5 | U.S. Geological Survey Land Use and Land Cover Classification System | 3-31 |
| 3-6 | Compatibility Between the U.S. Geological Survey and GEMLUP Land Use Classification Systems | 3-32 |
| 3-7 | Summary of GEMLUP Case Study Transportation Data | 3-34 |
| 4-1a | Instrumental Variables Used In The Initial Two-Stage Least Squares Analysis | 4-4 |
| 4-1b | Instrumental Variables Used In The Final Two-Stage Least Squares Analysis | 4-4 |
| 4-2 | Confirmation of Sufficient Identification In The Initial Two-Stage Least Squares Analysis | 4-5 |

LIST OF TABLES (CONTINUED)

| Table | Title | Page |
|---|---|---|
| 4-3 | Independent Variables Not Significantly Correlated With Dependent Variables In The Simultaneous Block of Equations . . . . . . . . . . . . . . . . . . . . . . . . . . . | 4-9 |
| 4-4 | Extreme Values Which Accounted for a Substantial Portion of Unexplained Variance in the Path Analysis and Produced Large Residuals . . . . . . . . . . . . . . . . . . . . . . . . | 4-13 |
| 4-5 | Comparison of Coefficients of Variation for a Pure Rates and a Mixed Totals-Rates Model . . . . . . . . . . . . . . | 4-25 |
| 4-6 | Comparison of the Number of Standardized Residuals Greater than 1.0 in Absolute Value For Two Model Forms . | 4-26 |
| 4-7 | Comparison of Path Regression Coefficients from the Final Path Analysis and the Jackknifing Stability Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . | 4-28 |
| 4-8 | Extreme Value Index of Stability for all Path Regression Coefficients in the Causal Model Equations . . . . . . . | 4-32 |
| 4-9 | Stability Index for all Path Regression Coefficients in the Causal Model Equations Based on the Effect of Variation in the Coefficients on the Dependent Variable . . . | 4-35 |
| 4-10 | Summary of Direct and Indirect Causal Effects on the Variable RES . . . . . . . . . . . . . . . . . . . . . . . . | 4-39 |
| 4-11 | Summary of Direct and Indirect Causal Effects on the Variable COMM . . . . . . . . . . . . . . . . . . . . . . . | 4-40 |
| 4-12 | Summary of Direct and Indirect Causal Effects on the Variable OFFICE . . . . . . . . . . . . . . . . . . . . . . | 4-41 |
| 4-13 | Summary of Direct and Indirect Causal Effects on the Variable MANF . . . . . . . . . . . . . . . . . . . . . . . | 4-42 |
| 4-14 | Summary of Direct and Indirect Causal Effects on the Variable HIWAYS . . . . . . . . . . . . . . . . . . . . . . | 4-43 |
| 4-15 | Summary of Direct and Indirect Causal Effects on the Variable EDUC . . . . . . . . . . . . . . . . . . . . . . . | 4-44 |
| 4-16 | Summary of Direct and Indirect Causal Effects on the Variable REC . . . . . . . . . . . . . . . . . . . . . . . . | 4-45 |
| 4-17 | Summary of Direct and Indirect Causal Effects on the Variable WHOLE . . . . . . . . . . . . . . . . . . . . . . . | 4-46 |
| 4-18 | Summary of Direct and Indirect Causal Effects on the Variable OTHER . . . . . . . . . . . . . . . . . . . . . . . | 4-47 |

LIST OF TABLES (CONTINUED)

| Table | Title | Page |
|---|---|---|
| 4-19 | Compilation of Nationwide Data on the Relationship Between Trip Generation Rates and Quantities of Land Use | 4-53 |
| 4-20 | Validation Results for Trip Production Equations Developed for Pittsburgh, PA | 4-54 |
| 4-21 | Suggested Validation Criteria for Trip Generation Rates | 4-55 |
| 4-22 | Trip Length Data Distributions Compiled by the Federal Highway Administration | 4-57 |
| 4-23 | Actual and Predicted Trip Length Data in Miles | 4-59 |
| 4-24 | Comparison of Actual and Predicted Trip Generation Rates by Land Use | 4-65 |
| 4-25 | Comparison of Actual and Predicted Daily Trips in the Area of Analysis by Land Use and Purpose | 4-67 |
| 4-26 | Comparison of Actual and Predicted Total VMT for the Area of Analysis | 4-68 |

# LIST OF FIGURES AND EXHIBITS

| Figure | Title | Page |
|---|---|---|
| 1-1 | Typical Elements of a Sewage Collection, Treatment & Disposal System . . . . . . . . . . . . . . . . . . . . . . . . | 1-5 |
| 1-2 | Technical Approach to the Development of a Statistical Model for Predicting the Growth Effects of Major Wastewater Projects . . . . . . . . . . . . . . . . . . . . . . | 1-16 |
| 2-1 | Example of the Displaced Growth Phenomenon Within a Hypothetical Drainage Basin . . . . . . . . . . . . . . . . | 2-18 |
| 2-2 | Hypothetical Legal Service Area of a Wastewater Major Project Illustrating the Locational Effects of the Project's Collection Network . . . . . . . . . . . . . . | 2-26 |
| 2-3 | Initial Specification of the Path Analytic Land Use Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . | 2-46 |
| 2-4 | Seven Equation Simultaneous Block in the Initial Path Analytic Land Use Model . . . . . . . . . . . . . . . . . . . | 2-47 |
| 2-5 | Two Equation Recursive Block in the Initial Path Analytic Land Use Model . . . . . . . . . . . . . . . . . . . | 2-48 |
| 4-1 | Final Path Analytic Model . . . . . . . . . . . . . . . . . . | 4-12 |
| 4-2 | Scattergram of Actual Versus Predicted Work Trip Lengths . . . . . . . . . . . . . . . . . . . . . . . . . . . . | 4-60 |
| 4-3 | Scattergram of Actual Versus Predicted Other Trip Lengths . . . . . . . . . . . . . . . . . . . . . . . . . . . . | 4-61 |
| 4-4 | Scattergram of Actual Work Trip Lengths Versus Model Residuals . . . . . . . . . . . . . . . . . . . . . . . . . . | 4-62 |
| 4-5 | Scattergram of Actual Other Trip Lengths Versus Model Residuals . . . . . . . . . . . . . . . . . . . . . . . . . . | 4-63 |
| 4-6 | Scattergram of Actual Versus Predicted VMT . . . . . . . . | 4-70 |
| 4-7 | Scattergram of Actual VMT Versus Model Residuals . . . . | 4-71 |

| Exhibit | Title | Page |
|---|---|---|
| 3-1 | Regional Planning Agency Survey Form . . . . . . . . . . . | 3-19 |

# I. INTRODUCTION

## A. STUDY OVERVIEW*

Pursuant to 40 CFR 51.12(d)-(h), State Implementation Plans must contain provisions to prevent any national ambient air quality standards from being exceeded. These provisions are called Air Quality Maintenance Area (AQMA) plans, and estimating the air quality impact of major land use and urban development projects is a necessary part of AQMA planning [42-47]. In addition, the National Environmental Policy Act [48] and the Council on Environmental Quality(CEQ) "Guidelines on the Preparation of Environmental Impact Statements" [49]** require the consideration of secondary impacts from major projects. CEQ states that:

> "Many federal actions, in particular those that involve the construction or licensing of infrastructure investments (e.g., highways, airports, sewer systems, water resource projects, etc.), stimulate or induce secondary effects in the form of associated investments and changed patterns of social and economic activities. Such secondary effects through their impacts on existing community facilities and activities, or through changes in natural conditions, may often be more substantial than the primary effects of the original action itself [49]."

This has been a particular concern for wastewater systems, since their primary impacts tend to be small, i.e., sewers and treatment plants generally improve water quality, but they may lead to significant negative secondary impacts. The probable large indirect impacts (redirecting growth and inducing development) of a new or expanded regional sewage treatment facility on ambient air quality, and the need for some procedure to ascertain what its impact will be, is recognized in the AQMA Guideline series ([42]:A-7ff, [45]:21ff). To date, EPA has not developed a model to estimate what the ambient impacts will be for use in AQMA planning. Thus, it was the purpose of this study to develop such a model.

This study is entitled the Growth Effects of Major Land Use Projects (GEMLUP-II), and it addresses the induced growth effects of wastewater major projects. Similar research on two other major project types, large residential developments and large industrial/office parks, has been performed previously and is reported on in the three volume set of GEMLUP-I final reports [17-19]. The secondary air quality impacts of wastewater projects are determined by the land use growth induced by such projects. That is, the air impacts are emissions from (1) residential complexes that appear at the end of and along new sewer lines, (2) other service-oriented land uses (commercial, industrial, office, government) that relate to residential development, and (3) motor vehicles used as transportation between development areas. Thus, the key to understanding secondary air quality impacts is to first understand the growth effects of a wastewater facility on land use in a region. The objectives of the study effort were: (1) to develop and test a path-analytic causal model that represents the induced land use ten years after construction of a wastewater major project, (2) to develop and validate a simplified predictive model of induced development, (3) to test and correct the GEMLUP-I VMT model [19], and (4) to develop worksheets that can be used to predict induced land use and associated air pollution emissions. For these purposes, data were collected from forty (40) case study wastewater projects nationwide.

---

* This section is partially based on material in reference [21].

** CEQ is in the process of promulgating regulations for the preparation of EISs.

The study project was divided into four major phases:

I. Definition of basic concepts and initial model specification,
II. Data collection,
III. Causal analysis of the land use model using path analysis,
IV. Development of predictive equations for the land use model and worksheet procedures.

Two separate technical reports are to be prepared. This is the first volume report and covers phases I-III of the study. The remainder of this chapter examines the technical background to induced development, gives a statement of the problem, an outline of the general study approach, and a glossary of terms used in this report. Chapters II-IV summarize the technical performance on phases I through III, respectively, while Chapter V summarizes the results and conclusions for these first three phases.

## B. TECHNICAL BACKGROUND

### 1. Federal Wastewater Funding Programs

Understanding the past history of federally funded wastewater collection and treatment projects is important to the current study since the identification of wastewater case study projects will depend to a large extent upon a review of published federal reports, as well as state and water/sewer district reports. Federal financial aid in the construction of municipal sewage treatment works was first authorized in 1948 (PL 80-845). This loan program was never implemented because necessary funds were not appropriated by Congress. The Federal Water Pollution Control Act of 1956 (PL 84-660) included the first authorization for federal grants to assist in the construction of waste treatment works. The Act authorized an appropriation of $50 million per year for such grants to be allocated to the States on the basis of relative population and per capita income. The grants were limited to 30% of the eligible project cost to a maximum of $250,000. Appropriations were increased during the early 1960's, and major amendments to PL 84-660 occurred in 1966. At that time, appropriation authorizations were increased, the maximum dollar limitation on grants was dropped, and the federal share was increased to a maximum of 55%. Most recently, the Federal Water Pollution Control Act Amendments of 1972 (PL 92-500) set up a three-phased program for the elimination by 1985 of pollutant discharges into the nation's navigable waters. First, an extensive matching grant program was instituted for the construction of municipal wastewater facilities (Sections 201-207). Under this program, the federal share has increased to 75% of eligible costs and annual expenditures have risen above $1 billion. Secondly, the National Pollution Discharge Elimination System has been adopted to control point source discharges through a permit program (Section 402). Finally, to coordinate the activities of the first two programs, an interrelated group of water quality planning processes are being implemented, viz. facility, area-wide, and state-wide planning (Sections 201, 208, 211, and 303). This coordinated effort of wastewater planning represents the first attempt by Congress to integrate land use planning with the control of water pollutant discharges.

Administration of this rapidly expanding federal program has passed through many different agencies over the years. Much of the earliest federal wastewater funding came from the U.S. Public Health Service (under HEW) and the Community Facilities Administration (the predecessor agency of HUD). From 1962 to 1965, federal funding of sewer projects was accomplished under the Accelerated Public Works Act of 1962 (PL 87-658) for which the Area Redevelopment Administration (ARA) of the Department of Commerce was the lead agency. In 1965, the Economic Development Administration took over these duties from the ARA, and, during the late 1960s the Appalachian Regional Commission also participated in the funding process. Finally, in 1970, EPA inherited most of the responsibility for overseeing all federal wastewater funding and planning.

## 2. Wastewater Facilities

### a. Basic Elements

Sewerage systems are required to collect and treat the materials entering them. The basic elements of wastewater management systems, shown in Figure 1-1, are the treatment plants which handle wastes delivered to them, and the various types of sewers carrying the material. The first stage of a sewer is the service connection to a residence, which transports waste from the household to small collector or lateral sewers serving streets. These feed into trunk sewers serving whole subdivisions. The larger sewers which receive wastes from a number of trunk lines are known generically as "interceptors", though this term originally meant only lines which "intercepted" untreated discharges to streams. The service area of a wastewater system is usually defined by gravity to include the watershed surrounding the lines, though use of pumping stations and force mains can allow the wastes to traverse watershed boundaries, and therefore extend the service area. A full discussion of the concept of legal service area is included in Section II.

Community wastewater facilities have generally been constructed in piecemeal fashion in the past. Some typical examples follow:

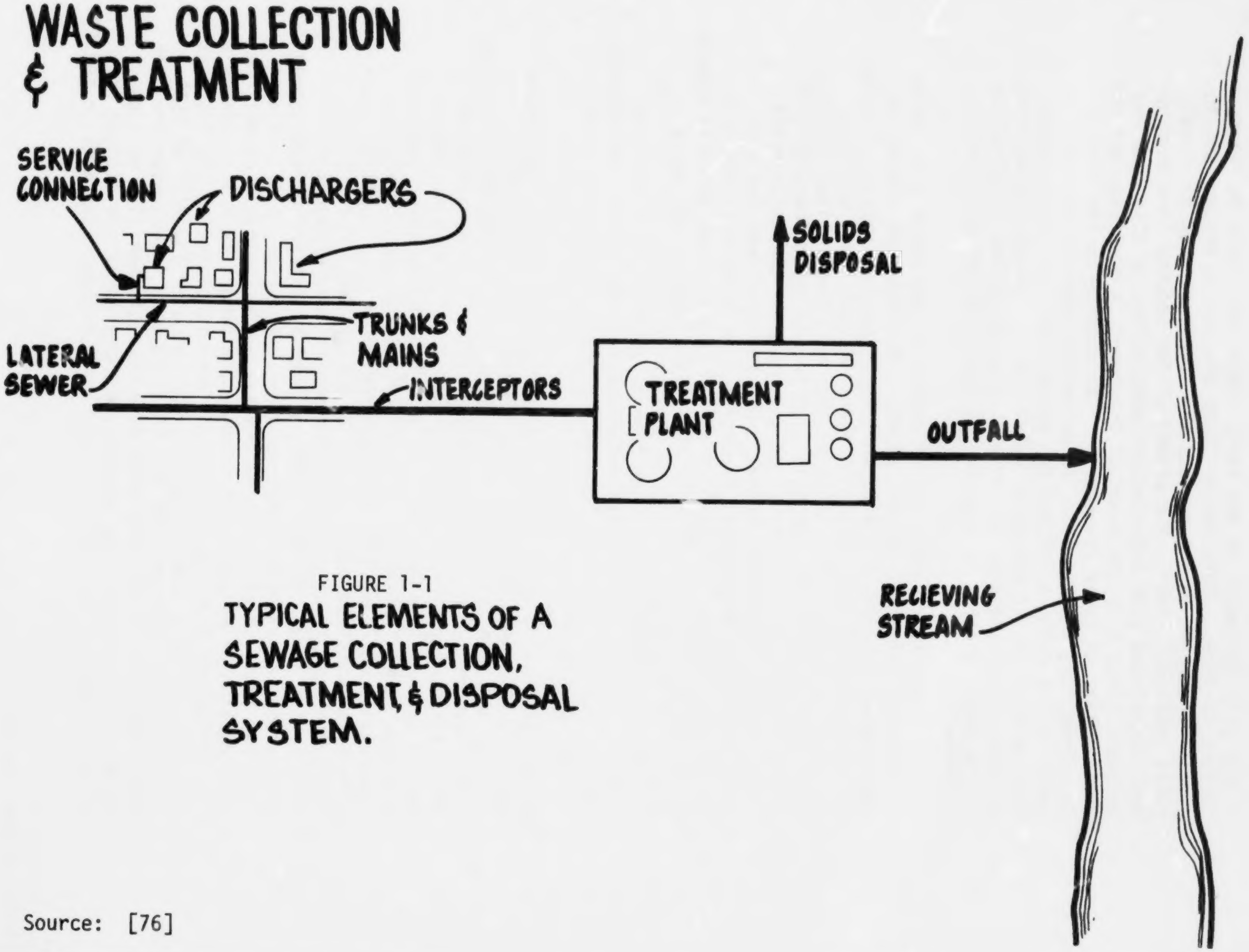

WASTE COLLECTION
& TREATMENT
SERVICE CONNECTION
DISCHARGERS
LATERAL SEWER
TRUNKS & MAINS
INTERCEPTORS
TREATMENT PLANT
SOLIDS DISPOSAL
OUTFALL
RECIEVING STREAM

FIGURE 1-1

TYPICAL ELEMENTS OF A SEWAGE COLLECTION, TREATMENT, & DISPOSAL SYSTEM.

Source: [76]

- A town puts in sewers in the early 1900s that drain to nearby streams or lakes. Interceptor lines are put in around 1960 to pickup discharges to the streams and transport them to a newly built treatment plant. The potential for induced land development associated with interceptor construction in this case may be small since only the means of treatment is changing, not the availability of sewer connections. An example is the *City of Los Altos, CA* [50]. In 1957, an extensive sewer system was built in Los Altos to pickup the waste load from septic tanks that were polluting the local water supply. Since the majority of land within the city was developed prior to the construction of the sewer system, no significant expansion has occurred subsequently.
- Several towns in a region put in sewers in the 1920s that drain to separate primary treatment plants. By 1960, the plants are used beyond capacity and equipment is in need of repair. Instead of constructing a new treatment plant in each town, one large regional treatment plant is built and interceptor lines (perhaps with pumping stations) are built out to adjoining towns to pick up each town's waste stream. In the process, the interceptor lines traverse unsewered, rural areas between towns and open this land up to development pressures. An example is the *City of Phoenix, AZ* [51]. In 1965, work was completed on the 91st Avenue treatment plant. This 45 MGD* facility picked up the waste streams of the nearby cities of Scottsdale, Tempe, and Mesa (in addition to Phoenix), allowing them to abandon their outdated plants. In the process, the interceptor lines crossed developable, vacant land in and around Phoenix, opening these lands up to intense development.
- A rapidly growing urban center updates its wastewater facilities by building a new treatment plant in 1960. Although sewer service has been available in the urban core for decades, outlying areas have relied on septic disposal systems. Due to the poor percolation characteristics of the local soil, however, these septic systems have been causing both health and water quality problems in the community. Therefore a large amount of excess capacity is designed into the new treatment plant to allow urban fringe and suburban areas access to the wastewater facilities. The extension of large interceptor sewers into predominantly rural areas opens up vast areas of land to intensive development pressures. An example is *Metropolitan Washington, D.C.* Between 1957 and 1966, 16 different wastewater projects involving the construction or extension of interceptor sewers in Prince George's and Montgomery Counties in Maryland were completed by the Washington Suburban Sanitary Commission at a cost of over $13 million [52]. More recently, between 1967 and 1973, 23 more interceptor projects were completed in the same area at a cost of over $170 million. As a result, Metropolitan Washington experienced the highest rate of growth during the 1960's of any large metropolitan area in the

---

*Million gallons per day.

U.S.[53]. Between 1960 and 1972, metropolitan population grew from 2 to 3 million persons. In general, this new development followed the interceptor lines into predominantly rural areas outside Washington, D.C. spreading suburban sprawl.

### b. System Capacity

The amount of sewage generated within a community can be simply defined as water consumption minus losses (e.g., lawn sprinkling) plus additions (e.g., infiltration). Sewage flow varies quite widely, showing seasonal, daily, and hourly fluctuations. For health reasons, wastewater collection systems are designed to accomodate the flow of the heaviest hour of the year for the design population. Therefore, interceptor lines are always built with "excess capacity" - that is, capacity beyond that needed to handle the peak flow from the present population of the area served. This excess capacity is available to developers to serve future projects that might connect to the lines. The proportion of excess capacity provided may vary from a low of 10-20% up to 200% or more, depending upon the growth implied by the design population forecast and the estimates of average flow and peaking factors. EPA recently reduced its design factors for sewers considerably because of the costs of overdesign.

### c. Package Plants

In the 1950s, in response to the demand for treatment facilities to serve isolated subdivisions and institutions, manufacturers developed miniaturized sewage treatment plants. Package plants can be regarded as filling the gap between individual septic tanks and large custom-designed municipal treatment plants[54], in the 1000 to 100,000 gallon per day range. For example, Morris County, NJ had 15 separate package plants in existence in 1965 to serve developing, suburban areas[55]. As interceptor lines have become available in this and many other areas, the waste load is diverted to sewers and the package plants removed.

### d. Wastewater Districts

Although the traditional organization to provide sanitary services has been the local government (municipal or county), regional commissions have also been formed for this purpose. Special districts, which have been created to operate a specific service and are largely financially independent, can be found today in all metropolitan areas immediately outside the central city and at numerous other locations within the U.S. serving small localities. The U.S. Census of Governments reports the existence of 1,411 special sewer districts in the U.S. in 1972 [56]. The direct reasons for their emergence has been the past refusal by the administrators of the existing central (municipal) systems to extend their service, such as sewer lines, very far into other political units, and the inability of the new suburban municipalities to finance a complete set of utilities. Federal incentives for area-wide approaches (under the Water Pollution Control Act Amendments of 1956 and 1972) have also increased the use of regional systems.

### 3. Land Use Impacts

The secondary effects of infrastructure investments are indirect long-term changes that occur as a result of operation or use of new facilities. Several recent studies have utilized an approach that analyzed the secondary impacts of wastewater facilities. A study examining the growth of northeast Philadelphia from 1945 to 1962 found a correlation between high density zoning and the enormous rise in real estate values of land with public sewer connections [57]. In metropolitan Washington D.C., Rabe and Hudson [58] report positive correlations between highway and sewer investments, and residential construction from 1960 to 1970, while Promise and Leiserson [53] have found statistically significant relationships between population growth, sewer capacity, and distance to interceptor lines. Shapiro and Tabors [59] developed estimates of sewer impact on land values for Clay, New York, also showing significant effects. A study of the influence of interceptor sewer policy on development in Fairfax County, Virginia, found that interceptor sewers affected the location of new sprawl of new subdivisions [60]. Based upon "self-fulfilling prophecies of inflated population projections" by county sewer authorities, sewer investments created "intense financial pressures for residential development". The rationale of the county authorities was that these major capital investments had to be repaid through "hook-up and service charge revenues on the service line" [61].

The CEQ has concluded that the secondary impacts of interceptor sewers upon undeveloped land at the urban fringe is the most critical element of waste treatment facility planning as related to land use planning [62]:

- "Interceptor sewers are defined as the major lines that run from the collector sewers to the treatment plant. Because the location of a new interceptor significantly increases the number of buildable lots along its right of way, a key issue is its capacity. There is a general tendency for such lines to be oversized in order to assure the necessary capacity for future development, but the oversizing itself can contribute to the extent of development that occurs."...
- "A related land use impact caused by large interceptor sewers is their tendency to be designed to run for long distances between existing towns before reaching the treatment plant. Such lines open up large areas of what may have been previously undeveloped land between the towns."...
- "Another phenomenon related to the construction of large interceptors is the tendency for developers to move immediately to the end of the new line in order to take advantage of both the available sewer service and the low land costs on the far urban fringe. The result is a costly leapfrog and fill-in development pattern, which increases the difficulty of properly planning the time and size of other public facilities and spreads the urban area out in a pattern that is wasteful of land and energy resources."

The significant relationship between extension of interceptor sewers into a community and the resultant urban sprawl patterns has been researched and confirmed in several CEQ-commissioned reports [61, 63-67]. While PL 92-500 and EPA regulations do not allow the use of federal funds for collector lines in sewer developments built after 1972 (Section 211), excess capacity for growth has been commonly included in the more major investments.

The costs associated with sewage disposal systems are not always realized at the time of installation. Builders prefer public sewered land for their developments since (1) the sewer connection fee is, in general, lower in cost to them than constructing individual septic systems, cesspools, or neighborhood package plants, and (2) they are not restricted by soil, drainage, or water quality restrictions in their development of the land. However, the true costs of a large wastewater collection and treatment system are borne by the residents of the area, not the developer, through increased property taxes to retire the large debt incurred. Thus, in The Costs of Sprawl [67], sprawl was determined to be the "most expensive form of residential development in terms of economic costs, environmental costs, natural resource consumption, and many types of personal cost."

In another report commissioned by CEQ [66], interceptor extensions were described as creating a direct incentive to the land developer to build at the urban fringe: "New sewers increase the density of possible development and thus the potential economic 'rent' (and development profit) per unit of land to the owner or developer." For example, a study of secondary impacts in New Jersey [80] concluded that:

> "...Sewers are the critical ingredient and the guiding force for growth in New Jersey. As the cost of land and construction rises, more townhouses and multi-family units will be built in proportion to single family homes. Sewers are essential for this higher density construction. As a result, the role of sewers as a growth determinant will become even stronger in the future."

In certain locations, sewers make medium and high density development feasible where previously, due to soil conditions and septic tank limitations, only single family homes on large lots could be built. Thus, adding sewers removes a development constraint. Their absence often prevents intense development of an area, although in the absence of laws prohibiting on-site disposal systems, not all development is precluded. Many times builders will develop an unsewered area, installing individual septic tanks without regard to their appropriateness, effectiveness, or water quality impacts. If these septic tanks do not function properly because of adverse soil conditiions, municipalities must rescue them by installing interceptor sewers. Thus, every homeowner pays double for sewer service. In addition, such improvement projects are far more costly than building sewers through open land due to the obstructions of existing structures, roadways, etc.

Attempts to control the land development associated with wastewater facilities have involved both delaying building permits and outright sewer moratoria. But recent studies of communities in which these policies are in use have shown that far from giving impetus to reconcentration of development in urban core areas, the moratoria can encourage sprawl by sending builders to jurisdictions farther out where there are no restrictions and land is cheap [68]. And even when such controls are effective, they can cause other problems in the areas of water quality and real estate values. An example community is San Diego, CA. This city currently exceeds the standards for several air pollutants, and like most of southern California, has a severe oxidant problem. San Diego County is also a high growth area due to its climate and large tracts of vacant, developable land. As a result, the region has had to rapidly expand wastewater facilities numerous times in the past. In the past few years, EPA has consciously delayed funding on construction of new wastewater facilities in the area due to the potential for their interfering (indirectly) with the attainment of air standards [77]. The effect of delays in construction funds has been widespread moratoria on connections to existing sewer lines and, thus, de facto moratoria on new residential construction. The net effects have been both good and bad. Uncontrolled growth has been halted for a while in San Diego, and the Comprehensive Planning Organization has been given significant powers for planning future development in the County. However, real estate values have skyrocketed to unrealistic values, speculation is rampant, and inadequate treatment of existing wastes, resulting from the lack of new treatment facilities, is causing serious water quality problems.

The above discussion of induced development represents a view which has also been stated as follows:

> "The comprehensive plan for an area represents a policy decision which is logically followed by implementation programs, one of which is the planning and construction of utilities. Precisely because there are no effective means to control overall "development" consistent with public goals, it is necessary and responsible to at least conform public spending for capital construction with adopted policy. Sewerage plans become 'self-fulfilling prophecies' by opening new areas for development, and therefore, often 'cause' the sprawl that functional planners claim to be simply serving after the fact" [69].

The planning community, however, is not unanimous in this view. For example, the director of New York's clean water program has the opposite viewpoint: "The granting of federal funds for pollution abatement facilities is not the cause of expanding land use, but is the cure for the effects of land use created by other sociological forces" [70]. Indeed, a 1965 survey of sewer service in the New York City metropolitan area found that, in general, sewers were lagging behind development in suburban areas, not causing it [71]. Package plants and septic systems were the principal means of wastewater treatment in the tri-state area in the 1950s and 1960s. Also, in Metropolitan Seattle, studies have shown that the implied cause-effect sequence (sewers "induce" developemnt) can run both ways. The Puget Sound Council of Governments has raised many alternative reasons for the perceived sewer-sprawl relationship [69]:

- "To what extent does sprawl merely infill behind the 1956 Federal Aid to Highways Act? Sewerage construction is funded on a high priority basis as the population withdraws from the central city. Is central city deterioration a cause of shifting "demand," or the consequence of construction-oriented categorical grants? Does the provision of utilities reinforce "demand" as well as merely responding to demand? What part of "demand" is due to constitutionally protected rights, and what part is due to real estate tax exemptions, capital gains benefits, average cost pricing of utilities, deductibility of taxes and interest paid on borrowed money, roads funded ironically on the basis of distance traveled (gas tax), etc."

Finally, a recent study for the National Water Commission [72] concluded that:

- "Federal investment in water resource development has not had noticeable effects on the growth or decline of communities. This assistance has for the most part come after a need is recognized, not before. While Federally supported water-related facilities may permit growth to occur, they do not cause its occurrence...

  ...Fundamental economic and location factors determine whether a community will grow or decline, and the availability of water-related facilities and services plays a minor role... (since)

  ...the costs for sewage collection and treatment represent a relatively small proportion of community intrastructure and service costs."

In summary, although land use patterns, and particularly the phenomenon of suburban sprawl, are the result of a complex set of historical, economic, social, and political interactions, the role of wastewater facilities in inducing development can be a major one. Thus, the growth effects of such projects should be recognized and dealt with to ensure the attainment of planning goals for a region, including the attainment and maintenance of air quality standards.

## C. PROBLEM STATEMENT

Extensive research has been done to design models for simulating the physical characteristics (demand, flow) of storm and wastewater systems. As a result, many dynamic models for designing entire networks are available (e.g., EPA's Storm Water Management Model (SWMM), Illinois' Storm Water Sewer Simulation Model (ISS), and the Systems Operational Analysis Model (SOAM)). Although research has also been undertaken to examine the secondary impacts of wastewater facilities (as discussed in previous sections), similar model development has not occurred. Predictive models of secondary development are required to answer questions concerning the impact, a priori, of infrastructure investments. One model that includes infrastructure as a determinant of land use growth is the EMPIRIC model [73] (or in its more recent form, PLUM), an "activity allocation model" that apportions population and employment projections to areas based on measures, in part, of sewer service availability. However, EMPIRIC does not simulate the more indirect long-term impacts of infrastructure investments, nor does it attempt to include any causal specification. Recently, research has been done to develop methodologies for secondary air quality impacts [78,79]. However, these approaches have proven to be less than innovative, since they generally use the design population projection for a sewage treatment plant as the basis for simple growth effects assessments.

Land use patterns and their attendant activities have a significant impact upon the type and amount of air pollution generated over a region. To the extent that land use can be associated with the discharge of pollutants, it is necessary to plan future land use and transportation that is compatible with acceptable levels of air quality. Comprehensive planning has in the past been relatively insensitive to air quality considerations. New air quality management procedures require that the planning community generate and use analytical tools to incorporate air quality constraints into the planning process [42-47]. A recent review of the state-of-the-art in models relating land use planning to air quality considerations [44] concluded that existing techniques are limited in application by the area-specific data base on which they are developed and their inability to disaggregate projected pollutant emissions by land use category. Thus, there is a need for a *generalized* analytical tool that can (1) predict land use development induced by major projects (such as wastewater facilities) by category type, and (2) convert such detailed land use projects into pollutant emissions for use in air quality management analyses. It was for these reasons that the GEMLUP-I model [17-19] was developed and the current study was undertaken to develop a model of the induced development associated with major wastewater collection and treatment systems. GEMLUP relates to a number of EPA programs [74], including AQMA planning, Environmental Impact Statement (EIS) review, the prevention of significant air quality deterioration or non-degradation, and the recent emissions offset policy promulgated for growth in non-attainment areas. Explicit or implicit in these programs is an evaluation of air quality impacts of land use plans or project developments. To perform such assessments, EPA has recently ruled that a portion of Section 208 grant money for waste quality management planning may be used for "air quality assessments of existing or projected development to be served by wastewater treatment facilities" [75]. Most recently**, Congress has amended the Clean Air Act, Section 316 of which now empowers the EPA Administrator to deny construction funds for wastewater facilities if an area is not covered by an SIP which "expressly quantifies and provides for" emissions related to secondary growth effects.

## D. GENERAL APPROACH

The principal objectives of this study were to formulate a statistical methodology to predict air pollutant emissions from:

- Induced land development associated with the construction and operation of wastewater facilities in a community.*

---

[*] The causal model of induced development used in this study does not take into account the effects of mitigating measures. Variables measuring restriction on on-site disposal and hook-ups to existing interceptor lines were included in the development of the model, however, see Section II.

[**] The EPA office of Transportation and Land Use Policy, Washington, DC, is in the process of preparing Guidelines for reviewing and mitigating secondary impacts of wastewater treatment facilities. These Guidelines will be available in early 1979.

- Motor vehicular traffic associated with the induced land development.

The ability to accurately predict the secondary development induced by major wastewater collection and treatment projects is dependent on understanding the complex interrelationships inherent in such a model. Thus, an important objective was to formulate and test a causal theory of induced development using path analysis. Additional objectives were to test and correct the motor vehicle traffic model developed as part of the previous GEMLUP-I study [19] for use in the air pollutant emissions projection procedure, and to integrate the predictive land use model, the traffic model, and land use based emission factors in a set of easy-to-use worksheets.

The approach to fulfilling these objectives involved the sequential execution of four separate project phases shown schematically in Figure 1-2 and summarized below.

In Phase I, our primary interest was to define the basic concepts on which this study is based. The first step in this process was to determine the modeling approach. Next, the infrastructure causal relationships of wastewater facilities in communities were studied and the knowledge used to define "induced development" in the model. The concepts of a "wastewater major project" and the "area of analysis" for induced development were also studied and defined. Knowing the important causal relationships enabled us to develop a list of model variables representative of the relevant factors involved (e.g., the major project, land use, regional growth). Finally, specific causal relationships between the variables were hypothesized and formalized in an initial causal model.

The principal objective of Phase II was to collect a sufficiently large, diverse, and thereby representative, cross-sectional data base on which to develop the model. Due to the critical importance of the quality of this data base and the large manpower effort required to collect such data, the first task required a careful selection of 40 case study wastewater major projects distributed on a nationwide basis, which had the potential, upon construction, to induce a significant quantity of land development in their communities and for which all the requisite data were available. To this end,

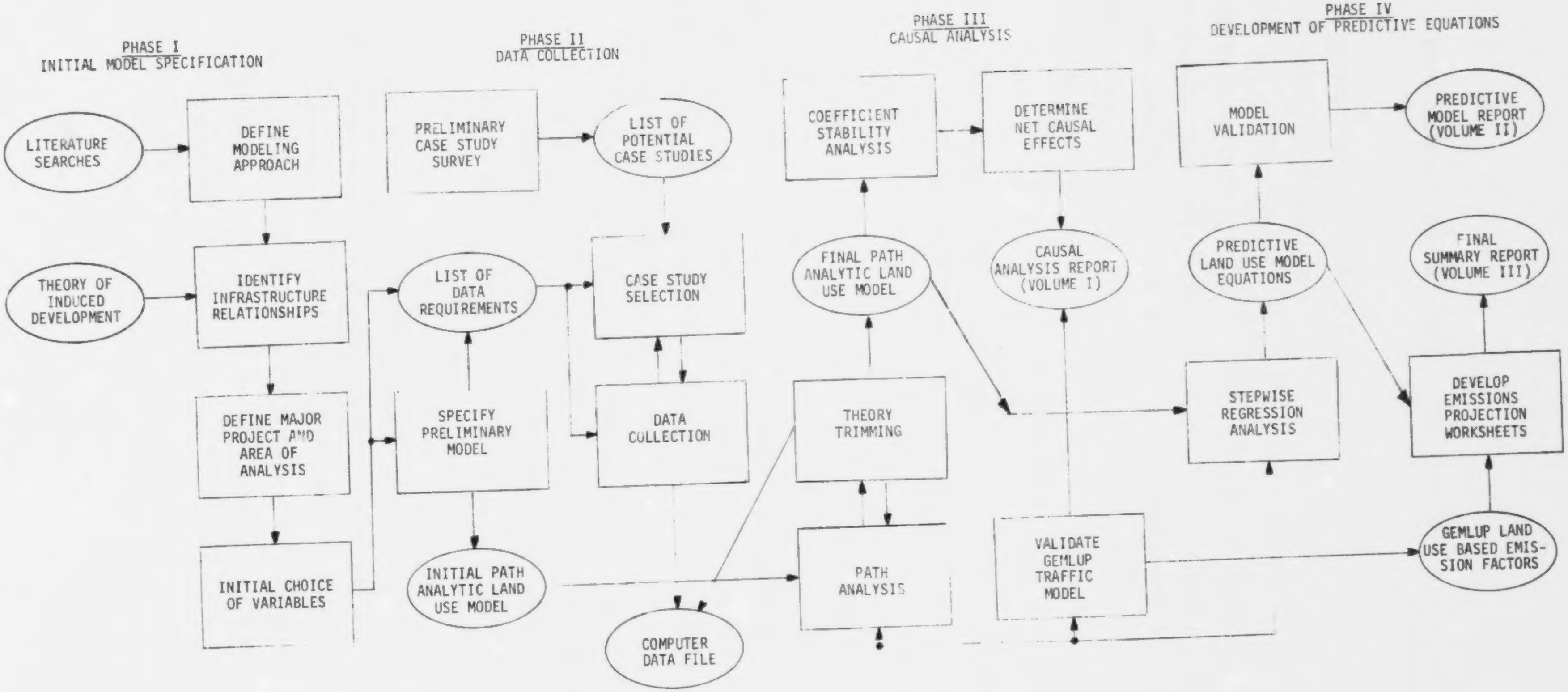

PHASE I
INITIAL MODEL SPECIFICATION
PHASE II
DATA COLLECTION
PHASE III
CAUSAL ANALYSIS
PHASE IV
DEVELOPMENT OF PREDICTIVE EQUATIONS
LITERATURE SEARCHES
DEFINE MODELING APPROACH
THEORY OF INDUCED DEVELOPMENT
IDENTIFY INFRASTRUCTURE RELATIONSHIPS
DEFINE MAJOR PROJECT AND AREA OF ANALYSIS
INITIAL CHOICE OF VARIABLES
PRELIMINARY CASE STUDY SURVEY
LIST OF POTENTIAL CASE STUDIES
LIST OF DATA REQUIREMENTS
CASE STUDY SELECTION
SPECIFY PRELIMINARY MODEL
DATA COLLECTION
INITIAL PATH ANALYTIC LAND USE MODEL
COMPUTER DATA FILE
COEFFICIENT STABILITY ANALYSIS
DETERMINE NET CAUSAL EFFECTS
FINAL PATH ANALYTIC LAND USE MODEL
CAUSAL ANALYSIS REPORT (VOLUME I)
THEORY TRIMMING
PATH ANALYSIS
VALIDATE GEMLUP TRAFFIC MODEL
MODEL VALIDATION
PREDICTIVE MODEL REPORT (VOLUME II)
PREDICTIVE LAND USE MODEL EQUATIONS
FINAL SUMMARY REPORT (VOLUME III)
STEPWISE REGRESSION ANALYSIS
DEVELOP EMISSIONS PROJECTION WORKSHEETS
GEMLUP LAND USE BASED EMISSION FACTORS

FIGURE 1-2

TECHNICAL APPROACH TO THE DEVELOPMENT OF A STATISTICAL MODEL
FOR PREDICTING THE GROWTH EFFECTS OF MAJOR WASTEWATER PROJECTS

a case study mail survey based upon the screening of over 15,000 federally funded wastewater collection and treatment projects nationwide was performed. Final selection of the 40 case study projects was made, a data collection training course was conducted for field personnel, followed by site visits to the case study areas to collect the required data.

The objectives of Phase III were to develop the final causal model and validate the GEMLUP-I VMT model [19]. The initial causal model was tested using the set of case study data and the statistical techniques of path analysis. This approach verified which of the hypothesized causal relationships were significant, trimmed those that were not, and determined model parameters for the final causal model. Tasks were also performed to trace the direct and indirect effects the model variables have on one another and to validate and correct the GEMLUP-I VMT model.

In Phase IV, predictive equations of induced land use associated with wastewater major projects were developed and validated. These predictive equations, the GEMLUP traffic model, and GEMLUP land use based emission factors [18] were then used as the basis for an emissions projection procedure. The procedure was developed in workbook form to serve as a generalized analytical tool for use by planners and environmental engineers in predicting the induced land use and air pollutant impacts associated with major wastewater collections and treatment systems.

## E. GLOSSARY OF TERMS

A dependent variable is the output variable of a mathematical function, which has as its arguments one or more input or independent variables.

Given a set of interrelated variables, an exogenous variable is one whose value is determined by forces external to or outside the system of relationships. An endogenous variable is an internal variable of the system which is completely determined (i.e., caused) by one or more exogenous and/or endogenous variables. The relationships of any one endogenous variable can be translated into mathematical form by designating it as the dependent variable of a function whose independent variables are all related causal factors, whether they be endogenous or exogenous variables.

Instrumental variables are exogenous variables which are introduced into nonrecursive systems of causal relationships (i.e., those involving interactions between endogenous variables) to allow the estimation of structural coefficients. A detail list of conditions associated with instrumental variables is given in Heise [15].

A wastewater major project is defined as the construction or extension of interceptor or collector sewer lines during the period 1958 to 1963 in a community in the United States. If construction date information is not available, then a grant funding date in the period 1956-1960 is acceptable. The project had to affect an increase in absolute system collection capacity of 1 MGD or more, and had to cost a minimum of $200,000 to construct. Phased projects are considered if the first phase of construction on the collection network was complete within 1958-1963 and the last phase of construction was complete by 1965.

Induced development is defined as urban land use associated with, caused, stimulated, or allowed by, or located because of the construction and operation of a major project.

Secondary development is induced development which has a direct causal relationship with a major project.

Tertiary development is induced development which has a direct causal relationship with secondary development, and so is only indirectly related to a major project.

The drainage basin of a wastewater project is defined as the land area which drains, by gravity, to any point of the collection network of the major project. In the case of essentially flat terrain (e.g., river deltas), this area is restricted to the locus of points no greater than 1,000 feet from any point along the interceptor line of the major project.

The legal service area of a wastewater major project is defined as the drainage basin of the major project plus any additional areas connected to the collection network by pumping stations and force mains.

The area of analysis is defined as the legal service area of a wastewater major project in the base year. It must be a minimum size of 5,000 acres (or approximately 8 square miles), and contain significant amounts of vacant developable land, some of which must be more than 5,000 feet from the nearest interceptor line in the base year.

"201" refers to Section 201 of the Federal Water Pollution Control Act Amendments of 1972 (PL 92-500). Section 201 calls for detailed planning for the wastewater treatment facilities needed to achieve the goals of the Act.

"208" refers to Section 208 of the Federal Water Pollution Control Act Amendments of 1972 (PL 92-500). Section 208 provides for the designation of state and areawide agencies for the purpose of developing effective water quality management plans for areas that, because of "urban-industrial concentrations" or other factors, have "substantial water quality control problems." The approach is aimed at integrating controls over municipal and industrial wastewater, storm sewer runoff, nonpoint source pollutants and land use.

## II. PHASE I - INITIAL MODEL SPECIFICATION

The first phase of the GEMLUP-II study consisted of a series of initial analyses to define basic concepts, specify the initial causal model, and identify data requirements. Thus, this phase served as a general guide and basic framework for the entire study.

### A. PERFORM LITERATURE SEARCHES

As the first step in Phase I of this study, literature searches were performed to identify all past research material on results and technical approaches to:

- Causal analyses and predictive modeling of induced development, in general, and of wastewater facilities specifically
- Simplified VMT predictive modeling procedures
- Land use identification and categorization from aerial photography, and the accuracy and precision of such techniques

In addition, searches were made to identify potential case studies from technical reports and wastewater trade journal publications. This material was needed at different points throughout the study.

The key to a successful literature search is the "search strategy" employed. This term refers to the application of an accumulated knowledge of: (1) what data sources to use, such as published directories, bibliographies, indexes, computer data bases, and (2) how to use them, including the manipulation of Boolean logic in developing complex keyword descriptors for search purposes. Specific, effective search strategies were developed for the above purposes with the assistance of information specialists in Abcor's Technical Information Center (TIC), and these were applied both manually and using computerized retrieval systems. The data bases searched on are the following:

- Air Pollution Abstracts (APTIC)
- Engineering Index (COMPENDEX)
- Environment Abstracts (ENVIROLINE)

- Geo Abstracts F: Urban and Regional Planning
- Journal Indexes:
  - ASCE Journal of Surveying and Mapping
  - ASCE Journal of Urban Planning and Development
  - Environment and Behavior
  - Environment and Planning
  - Transportation Engineering
- Monthly Catalog of U.S. Government Publications (MOCAT)
- National Technical Information Service (NTIS)
- Pollution Abstracts (POLLUTION)
- Science Citation Index (SCISEARCH)
- Smithsonian Science Information Exchange (SSIE)
- Urban Affairs Abstracts
- Water Resources Abstracts (WRSIC)

All relevant citations obtained are listed in Appendix D by search objectives, along with references obtained from past work in these technical areas. Thus, Appendix D serves as a comprehensive bibliography for this study.

## B. AERIAL PHOTOGRAPHIC INFORMATION

Since aerial photographs were the principal source for land use data in this study, a task was performed to determine the typical errors associated with land use data estimated in this manner. Appendix C provides a complete summary of this task.

## C. DEFINE MODELING APPROACH

### 1. Theory of Induced Development*

In order to develop a causal model of the induced development associated with a wastewater facility, it was necessary to first hypothesize a theory of induced development. Of the many types of scientific theory discussed in the literature [22,23,81-83], the most rigorous type of theory that we could ascribe to is "factor theory". Factor theory involves a selective, explicit enumeration of all factors thought to influence a given phenomenon (in this case, land use), is characterized by narrow and non-overlapping generalizations, and utilizes empirically defined variables to represent the factors involved. While almost every effort at causal explanation involves factor theory, it is limited theory because it does not readily suggest other generalizations due to its relative narrow focus. Consequently, the factor theory was operationalized in the form of a model.

---

*This section and the next are based in part on material written by Thomas McCurdy, Project Officer, in February 1976.

The preliminary hypothesis of induced land use development was an elaboration of the following theory:*

- Construction of a regional wastewater treatment facility enables interceptor sewer lines to be extended into previously unsewered areas. The availability of public-financed sewer connections on vacant land is one factor that encourages large industrial and/or residential developments; industrial complexes generate jobs which result in the construction of additional nearby residential development, and all of these induce retail establishments to locate near them and generate demand for community facilities. Increased development leads to the construction of streets and highways to improve accessibility to the area. Better access fosters continued urban development, particularly highway-oriented industrial, office, and commercial land uses. Additional sources of employment spur on another round of residential development, and so forth.

## 2. Approach to Theory Testing

Selection of an approach for testing the theory of induced development was guided by the study objectives and constrained by resource limitations. Since the theory was only a preliminary hypothesis of the true causal relationships, a deterministic approach** was inappropriate due to its dependence on the specification, *a priori*, of the correct model structure. That is, use of a deterministic model would not allow one to test or refine the theory of induced development. Also, since it was desired to have a model which would represent the average cause and effect response between variables, it was necessary that the approach chosen be capable of dealing with the random errors present in case study data. A deterministic model does not recognize such errors; therefore, it was decided to use a statistical approach.

---

* The theory is not new. Most of the urban development models referenced later in the text are based upon the same general relationships posited in our theory; they are usually less explicit than the GEMLUP model, however.

**For example, see Lowry [84], Forrester [85], Hill [73], Seidman [86], and Center for Real Estate and Urban Economics [87]. In addition to these general models there are many single sector models that have been developed since the early 1960's. See references [44] and [88] for a review of urban development models.

Limitations of time and resources made a dynamic modeling approach infeasible, because of the effort involved in obtaining sufficient longitudinal data to incorporate time into the system. A difference approach would involve solving a system of partial differential equations for the change in development patterns over time. If used, however, the solution would be deterministic and therefore inappropriate, as previously discussed. Since the study objective of quantifying causal relationships required the use of a technique which could represent the direction of causal action between an ordered set of model variables, a static approach using path analytic techniques (based on regression analysis) was selected to test the theory of induced development. This approach involved formalizing the theory in a causal path diagram and then testing it using cross-sectional data and the statistical techniques of path analysis. Mathematically, this is equivalent to specifying and solving a system of simultaneous, linear equations in which some of the sources of variations (i.e., unobserved variables) may be unspecified and where the number of observations exceeds the number of variables. Three problems associated with unobserved variables can arise with this type of approach. First, there may be important causal variables which are left out of the model. Second, the variables chosen may not be true measures of the effects being researched. Third, there will always be random effects (residuals) in any system which cannot be modeled. In response to the first problem, we specified as complete an initial model hypothesis as possible to avoid leaving out any significant variables. As a result, we assumed all other unobserved effects to be random in nature. We did not address the second problem because the sample size placed severe constraints on the depth of our analysis into the model structure. The third problem is dealt with through the assumption that we are dealing with random effects (residuals) which are independent with each other, and with the exogenous variables in the model. Thus, their covariances are not needed or included in the model. Path analysis is not restricted by this assumption; it is adopted only heuristically to keep the model form simple enough for solution. The assumption is justified to the extent that all major system variables have been specified explicitly in the model [15].

The static approach to testing the inherently change-oriented theory was justified by three factors: (1) the theoretical assumption that induced development follows a single basic causal structure for all cases or observations, (2) the use of a cross-sectional data for variables observed in a static state and the assumption that input variables are initialized at time zero and held constant long enough so that all the causal consequences in the system are realized, and (3) the use of time-lagged exogenous variables in the system. The conceptual usefulness of these factors in testing causal theory is well described in Heise [15] and Blalock [89].

### 3. Model Form and Characteristics

A thorough review of the technical literature did not provide enough information on induced development associated with wastewater facilities to be able to pre-define the form of relationships in the system. The form, therefore, was assumed to be linear. This is not a bad *a priori* approximation in most social science applications, and it allows the use of well-developed statistical techniques to test for causality. There is accumulating evidence that many social systems can be approximated by a linear function as long as operating conditions remain fairly stable [15]. Even complex nonlinear relationships can often be approximated by a constant relation in discrete subregions of the relationship [90]. Also, if a relationship is thought to be nonlinear on theoretical grounds (e.g., multiplicative, exponential), the data can be transformed prior to entering it in the linear analysis.*

A time period of 10 years, specifically 1960 to 1970, was used in the land use model. The selection of this time period is based on two factors. First, from a preliminary review of the technical literature related to this study, it was judged that a time period of 5-10 years was reasonable for allowing the secondary and tertiary impacts associated with a wastewater majo project to be realized. Second, the principal source for the comprehensive demographic and economic data required in the model was the U.S. Census, taken at the beginning of each decade. Although some states undertake a census at mid-decade, it was determined that these efforts are limited in scope and sporadic, and do not provide either the detail or quality of data available from the U.S. Bureau of Census [91].

---

*As part of the testing of the path analytic model, nonlinear forms were considered.

The possibility of using the GEMLUP-I model [19] as a portion of the GEMLUP-II model was examined. It was found, however, that such a "piggy-backed" model would be unworkable since the GEMLUP-I models were developed for a fixed area of analysis while the GEMLUP-II model will use a variable area of analysis.

In summary, the GEMLUP causal approach consisted of testing a cross-sectional, path analytic (simultaneous equations) model that explains total land use development in the vicinity of a wastewater major project 10 years after its construction.

### 4. Technique Selection for Path Analysis

Two statistical techniques were used to verify the hypothesized path analytic model of induced land use and to determine model parameters: two-stage least squares and ordinary least squares multiple regression. The first technique was used to solve for path coefficients in the model equations affected by simultaneity, i.e., those containing feedback causality between variables. Ordinary least squares regression was used to solve for path coefficients in recursive (not interconnected*) equations. There are many other multivariate linear statistical techniques available for data analysis purposes. Presented below are brief discussions of why these other techniques were not used in the current study.

#### a. Factor Analysis and Canonical Correlation

Unlike forms of regression analysis, these techniques do not make a distinction between endogenous and exogenous variables [92]. Thus, the direction of causal action, an important element of this study, cannot be specified. Also, if a simultaneous treatment of all causal factors is to be considered with these techniques, then the individual relationships between specific endogenous and sets of exogenous variables, which are desired, cannot be obtained. Finally, the factors (e:g., principal components) and canonical variates output by these procedures can defy description in other than mathematical terms, i.e., they cannot be related to measurable, physical quantities. Thus, a theory of causality cannot be verified with these techniques.

---

*And assuming independent errors.

### b. Three-Stage Least Squares and Maximum Likelihood Estimation Techniques

Three-stage least squares is an alternative path analytic technique which involves essentially performing a two-stage least square analysis, obtaining the residual errors, and then *simultaneously* re-estimating the model coefficients in all model equations. Three-stage least squares is called a *full-information method* because of the last stage in which all the structural coefficients are determined at one time. The advantage of this approach is that it produces more efficient (i.e., precise) model coefficients, although not necessarily more accurate ones. It is a tool for fine tuning the coefficients in a model. The principal disadvantage of any full-information method is that it incorporates model specification errors of any one equation into the calculations for all the model equations [15]. Thus, if but one of the equations is mis-specified, the errors involved may cascade throughout the entire model. For this reason, these methods are most appropriately used when there is a high level of confidence in the theoretical specification of causal linkages. A second disadvantage to full-information methods is the reduction in the available degrees of freedom for statistical fit due to the simultaneous consideration of all model variables in one estimation step. Due to the distinct possibility for error propagation in using three-stage least squares, and the problems associated with small sample size, this technique was judged inappropriate in the current study.

The maximum likelihood technique is an alternative method for estimating the coefficients of the path analytic model, which differs from two-stage least squares in that it offers the ability to specify the form of the relationships between unobserved variables (not in the model) and those included in the model structure. The price for being able to make these new assumptions is increased model complexity and statistical constraints. The latter is particularly troublesome when the sample size is small, and in some instances it may be impossible to use such techniques without first discarding several model variables to free up more degrees of freedom in the statistical analysis. This technique represents a very general approach that allows one to specify a more complex model structure, perhaps involving correlations between error terms or causal relationships between observed and unobservable variables. This technique is most appropriate when the causes and effects being researched are not directly measurable and it is necessary to represent them by related indicator variables which can be observed. In the current study, indicator variables are not a problem since the principal model outputs, the endogenous land use variables, are specifically what we wish to study (and ultimately predict). Also, there is no basis for assuming a more complex model structure in the current study that would justify the use of these techniques. And if there were, it is doubtful the limited sample size could accommodate such complexity. In addition, maximum likelihood techniques require the assumption of multivariate normality (which two-stage least squares does not). Finally, these techniques are full-information methods, and so suffer from the same disadvantages as three-stage least squares. For the above reasons, these techniques were judged inappropriate for the current study.

### c. Robust Regression

Robust regression is a class of experimental statistical techniques that differ from ordinary least squares in certain aspects [93]. For example, robust regression may involve regression on the median rather than the mean of the data, or it may involve a weighted least squares where observations far from the mean are discounted. Other approaches are to eliminate so-called outlyers completely in the analysis or substitute in a fixed value. The principal advantage of robust regression is in the detection and modification of outlying data points. Implicit in its use is the assumption that such outlyers are undesirable and are introducing errors in the estimation of structural coefficients. In the current study, there is no basis for making such an assumption, and thus no basis for weighting observations. Due to the small sample size employed, an outlying data point may simply be the only value available in a particular portion of a variable's range. For this reason, robust regression was not used in this study.

### d. Ridge Regression

This technique is a modification of ordinary least squares in which a dampening factor "k" is included in the process of estimating model parameters [94-96]. The advantage of ridge regression is that it produces more stable estimates of the regression coefficients than ordinary least squares. In order to do this, the traditional assumption of unbiased estimates has to be discarded. Ideally, the ridge estimates will be closer to the true model parameters because the average instability error (imprecision) of ordinary least square estimates is greater than the bias (inaccuracy) of the ridge estimates. This will be the case if the "correct" value of k is chosen. Unfortunately, there is not established technique for choosing the best value of k, and in general, a trial and error approach is required.

Due to its capabilities, the ridge regression is designed to handle many of the problems inherent in the GEMLUP path analysis, such as unstable coefficients, small sample size and multicollinearity. For this reason, the technique was seriously considered for use in this study. Since it is not a replacement for two-stage least squares, but only a modification of ordinary least squares, it would have to be incorporated into both steps of two-stage least squares. A standard computer program for ridge regression does not exist*; therefore, extensive software development would have been required. Also, since the algebraic solution to two-stage least squares would no longer hold when the k factor was introduced, that analysis would have to be done as two ordinary least square analyses, with an interim manipulation of residuals. This would greatly increase data analysis costs. Finally, each equation would have to be solved numerous times, each with a different k value, the results plotted, and an optimum k chosen from the graphs. Limitations of time and resources precluded such a significant amount of additional work, and therefore ridge regression was not used in this study. This is unfortunate, since the problems associated with small sample size are the most serious limitations imposed on the path analysis by its use in the GEMLUP study.

---

*One is currently being developed as part of the SPSS [92] system, however, there is no projected date of availability.

### e. Modification of Two-Stage Least Squares for Undersized Samples

One major problem encountered in the first GEMLUP study was the problem of having more exogenous variables than case study samples. When this is the case, the sample is said to be undersized. The approach used previously was to retain for the analysis only those variables which theory indicated were important and those that correlated highly with the endogenous variables of the system. Thiel [97] suggests a modification of the two-stage least square technique that permits the retention of all the exogenous variables in the analysis. The technique is based on the asymptotic (large sample) properties of least squares analysis. While the technique does not appear to have been employed often it does have great promise and would be extremely useful in this study. However, since a standard computer program for Thiel's method does not exist, extensive software development would have been required. Limitations of time and resources unfortunately precluded the associated additional work.

### f. Other Considerations

In addition to the above techniques, several others were suggested to deal with problems that sometimes arise in statistical data analysis. The first of these problems is heteroskedasticity, which involves unequal variances in the model equation residuals. This problem was addressed in the path analytic work by normalization of all model variables, as described on pp. 2-14 and 2-15.

A second problem that was identified is intrinsic non-linearity in the causal relationships. Many techniques exist for dealing with this problem, however, they were generally not applicable to the current problem since we found no basis (through a review of the technical literature) for assuming non-linearity or a particular non-linear form. Thus, a linear form for the path analytic model was assumed, as previously discussed. The validity of this assumption was checked by considering other, non-linear forms in testing the model and comparing the results.

The related problems of autocorrelation and serial correlation in the case study data were mentioned, along with Box-Jenkins techniques. It was judged that these problems do not exist in this study since the independent variables employed in the model are defined for a single point in time, the base year of a major project, and so temporal correlation cannot occur between variables.

The problem of discontinuous dependent variables was raised and the techniques of logit analysis and discriminant analysis mentioned as possible solutions. It was judged that this issue was not a problem in the current study, since all of the dependent (land use) variables are by nature continuous in their function.

## D. DEFINE INDUCED DEVELOPMENT

Induced development is defined as urban land use associated with, caused, stimulated, or allowed by, or located because of the construction and operation of a wastewater facility. This definition provided a basis for the selection of endogenous variables for the causal analysis. The approach used to selecting these variables involved a review of the technical literature to identify what types of land use are principally induced by a wastewater major project (i.e., what are the infrastructure relationships of wastewater facilities in a community), and which are measurable in an objective way, for inclusion in the causal and predictive models. The initial choice of model variables was also constrained by the fact that model outputs must be compatible with the land use based emission factors [18] developed in the previous GEMLUP study, which were used in Phase IV of this study to produce an emissions projection procedure.

Several recent studies have examined the secondary impacts of wastewater facilities, and reached the following similar conclusions. Wastewater facilities have their greatest growth-inducing impacts on single family housing [58,60,61,64,66]. They also encourage medium to high density housing [57,58,61, 64,66] by removing the development constraint associated with soil conditions and septic tank limitations. Other important land uses for which a wastewater infrastructure acts as a stimulus are industrial (manufacturing), commercial, educational, transportation and recreational developments [53,61,66]. Due to the large effect of wastewater major projects on residential and industrial development, the situation occurs in which these secondary developments then induce other tertiary land uses, namely commercial, office-professional services, manufacturing, wholesale-warehousing, non-expressway highways, hotel-motel, cultural-entertainment, religious, educational and recreational developments [17]. Note that the implied temporal ordering here is not represented explicitly in the GEMLUP model; rather, the objective is to illustrate how a wastewater major project can induce growth of many different land use types. The initial list of endogenous variables, given in Table 2-1, reflects all of the land uses discussed above, which are thought to be ultimately induced by the construction and operation of wastewater projects. In addition, it includes a disaggregation of the residential and commercial land use categories into more specific sub-categories based on size. These detailed data are required to accurately estimate associated air pollutant emissions, through application of the GEMLUP land use based emission factors [18]. Also, this disaggregation will allow the user of the GEMLUP model to introduce a specific density distribution for residential and commercial land use that reflects local and future development trends, rather than the density patterns of the period 1960-1970. The endogenous variables listed in Table 2-1 will represent, in the GEMLUP model, predicted land use in the area of analysis 10 years after the construction of a wastewater major project.

Certain land use categories are not included in the initial list of endogenous variables. Some of these are independent of wastewater facilities, such as passive open space, agriculture and mining. Others, though requiring some form of wastewater services, are usually the result of decisions made by regional, state or federal governments, and so their construction and location is generally not determined by the existence of a wastewater major project.

TABLE 2-1

INITIAL LIST OF ENDOGENOUS VARIABLES

| | Land Use Type | Units |
|---|---|---|
| 1. | Residential<br>a. Single family detached (1 unit/structure).<br>b. Single family attached (2 units/structure).<br>c. Mobile homes.<br>d. Multifamily low rise (3-4 units/structure).<br>e. Multifamily high rise (5+ units/structure). | Dwelling Units |
| 2. | Commercial<br>a. <50,000 sq. ft. of gross leasable area (GLA)<br>b. 50,000-100,000 sq. ft. of GLA<br>c. >100,000 sq. ft. of GLA | 1,000 $ft^{2*}$ |
| 3. | Office | 1,000 $ft^{2*}$ |
| 4. | Manufacturing | 1,000 $ft^{2*}$ |
| 5. | Wholesale-Warehousing | 1,000 $ft^{2*}$ |
| 6. | Non-expressway highway | Lane Miles |
| 7. | Educational | 1,000 $ft^{2*}$ |
| 8. | Active, Outdoor Recreation | Acres |
| 9. | Other Urban Land Uses (hospitals, hotels and motels, cultural facilities, churches) | 1,000 $ft^{2*}$ |

* Of total floor area.

Examples of the latter are correctional institutions, military installations, universities, airports and sport stadiums or arenas. Package treatment plants are often used for these facilities, allowing practically any location where vacant land is available.

Certain land uses are combined in the initial list of endogenous variables. The category of "other urban land uses" includes variables which are only indirectly related to wastewater major projects, and which by themselves, involve very small amounts of land. Thus, on an individual basis, these categories are very difficult to predict. So, they are aggregated.

Another issue associated with definition of induced development was determining whether to model total land use, or change in land use, at the end of the 10-year period of analysis. It was decided to model total land use for three reasons. First, data were not available to efficiently determine changes in land use. 1970 land use was extracted from aerial photographs, USGS quadrangle topographical and land use/land cover maps, and local land use maps for that time period. All of these sources were generally less readily available for 1960. Second, the air quality emissions projection procedure required predictions of total land use as input variables. If only change in land use was predicted, a step would have to be included in the worksheet procedure to add in base year land use quantities. Since emission projections are the principal output of the GEMLUP tool, it is important to keep the computational procedure as simple as possible. Finally, if change in land use is desired, it can be obtained by subtracting base year amounts from model predictions. This information is less likely to be desired than estimates of total air pollutant emissions.

The endogenous variables listed in Table 2-1 were normalized prior to their use in the path analytic model. This requirement was due to the selection of a variable area of analysis which exhibited wide variability between case study major projects.* For example, consider the form of the GEMLUP model for a recursive equation:

$$Y = \beta_o + \beta_1 X_1 + \ldots\ldots + \beta_p X_p + \varepsilon \qquad (1)$$

---

*Exogenous model variables were specified on a density or percent basis, i.e., normalized as well, due to the variability in area of analysis. See page 2-31.

where Y is the endogenous variable; $X_1, \cdots\cdot, X_p$ are the exogenous variables; $\beta_0, \beta_1, \cdots\cdot, \beta_p$ are the model coefficients; and $\varepsilon$ is the error term. In estimating the $\beta$'s, it can be shown that if the standard deviation of $\varepsilon$ is constant for all case study samples, then ordinary least squares will produce good estimates of the $\beta$'s. However, if this is not the case, then adjustments are required before employing the least square technique. One such example of when an adjustment is needed arises in this study due to the wide variability in areas of analysis in which the endogenous variable Y is measured. As the total area increases, so will the magnitude of Y, in general, and thus so will the standard deviation pf $\varepsilon$. That is, large projects are apt to be more variable (i.e., larger standard deviations) than smaller ones regardless of the values of the X's. If this is the case, than a more appropriate model than (1) is

$$\frac{Y}{f(A)} = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \eta \qquad (2)$$

where f(A) represents an appropriate transformation of the area of analysis A so as to produce an error term $\eta$ with constant standard deviation. The function f(A) is determined as follows. For a fixed set of X's, if the standard deviation of Y ($\sigma_y$) is found to be proportional to A, then f(A) = A; if $\sigma_y$ is found to be proportional to $\sqrt{A}$, then $f(A) = \sqrt{A}$, and so forth [12]. Since a very large number of case study samples would be required to analyze the variability of ($\sigma_y$) with A, for a fixed set of X's, the technical literature was examined. No references could be found related to this question, however. Thus, the intuitive judgment was made that the standard deviation of Y is proportional to A directly, and so f(A) = A was used.

A separate problem associated with defining the induced development is the approach used to investigate the phenomenon of displaced growth. A wastewater major project will induce changes in land uses within its region by two mechanisms. First, the project will affect the growth of the region as a whole, and second, the project will produce changes in the pattern of growth within the region by a substitution or displacement effect. This latter effect is characterized by growth that is not "new" or directly induced by the project itself, but rather is a relocation within the region to the area of analysis as a result of the project. At a particular location in the region, the substitution effect of the project may be either positive or negative - the project may act to attract growth to the immediate area that would not have occurred there without the project, or lead to a relocation of the activity to another area in the region. Likewise, the major land use project could make the immediate area less attractive to certain land uses and, therefore, make other locations comparatively more attractive. To summarize, a major wastewater project induces changes (simultaneously) in both the *total* land use and its *location* within a region.

One way to categorize types of induced development is the following:

- Development that is new to a region, attracted by the major project from other regions
- Development that is indigenous to a region, but located in the specific area of analysis due to the major project
- Development that is "pent up" in the area of analysis and allowed to occur by the major project.

Another approach is that any development can be viewed as using up or *displacing* a development potential that might have been used elsewhere. Thus, development is never "indigenous" to, or "pent up" in an area, but rather development potential is directed to, or located in, areas most favorable to it. A more fundamental viewpoint reflects the fact that development is rarely caused by one key factor (wastewater facilities), but rather a combination of variables. Since it is virtually impossible to determine the exact source of development potential which manifests itself in the case study project areas, it was decided that the key to understanding and defining the induced growth of a major wastewater project is a methodology to measure and analyze displaced growth effects. Or stated differently, the need to address the issue of displaced growth is due to the fact that, even though the model of induced development produced by this study is project-specific, the effects of a wastewater collection system are often more regional in nature, attracting (i.e., displacing) growth from land areas surrounding the service area in which public sewer connections are most easily available. An illustration of this is given in Figure 2-1, where a hypothetical drainage basin is shown in which an urbanized core has been sewered for many years, although only primary treatment was applied. The construction of a new regional secondary treatment plant in a nearby large city allows an interceptor line to be built into the area to pick up its entire waste load. Assuming excess capacity has been built into this line, a new (publicly financed) collection system rapidly spreads through a section of the vacant land of the community, providing obvious development potential. This example illustrates that it is necessary to predict not only the total amount of future land use in a community, but where development will occur, e.g., in sewered or unsewered areas. New development will obviously not occur exclusively in vacant, sewered areas since other locational factors (e.g., transportation facilities, availability of markets, zoning) can be equally or more important. However the existence of sewer connections can be a deciding influence and, therefore, a cause of displaced growth.

A technique for estimating displaced growth effects will be outlined in the second volume technical report.

## E. DEFINE MAJOR PROJECT

For the purpose of model specification and case study selection, an explicit definition of what constitutes a major wastewater project was determined. This definition was based on the premise that we are only interested in wastewater projects which have the *potential* for inducing and/or redirecting land development, i.e., those that have the effect of increasing or expanding the number of sewage connections in a service area. Examples of projects that do not have the potential for significant induced growth

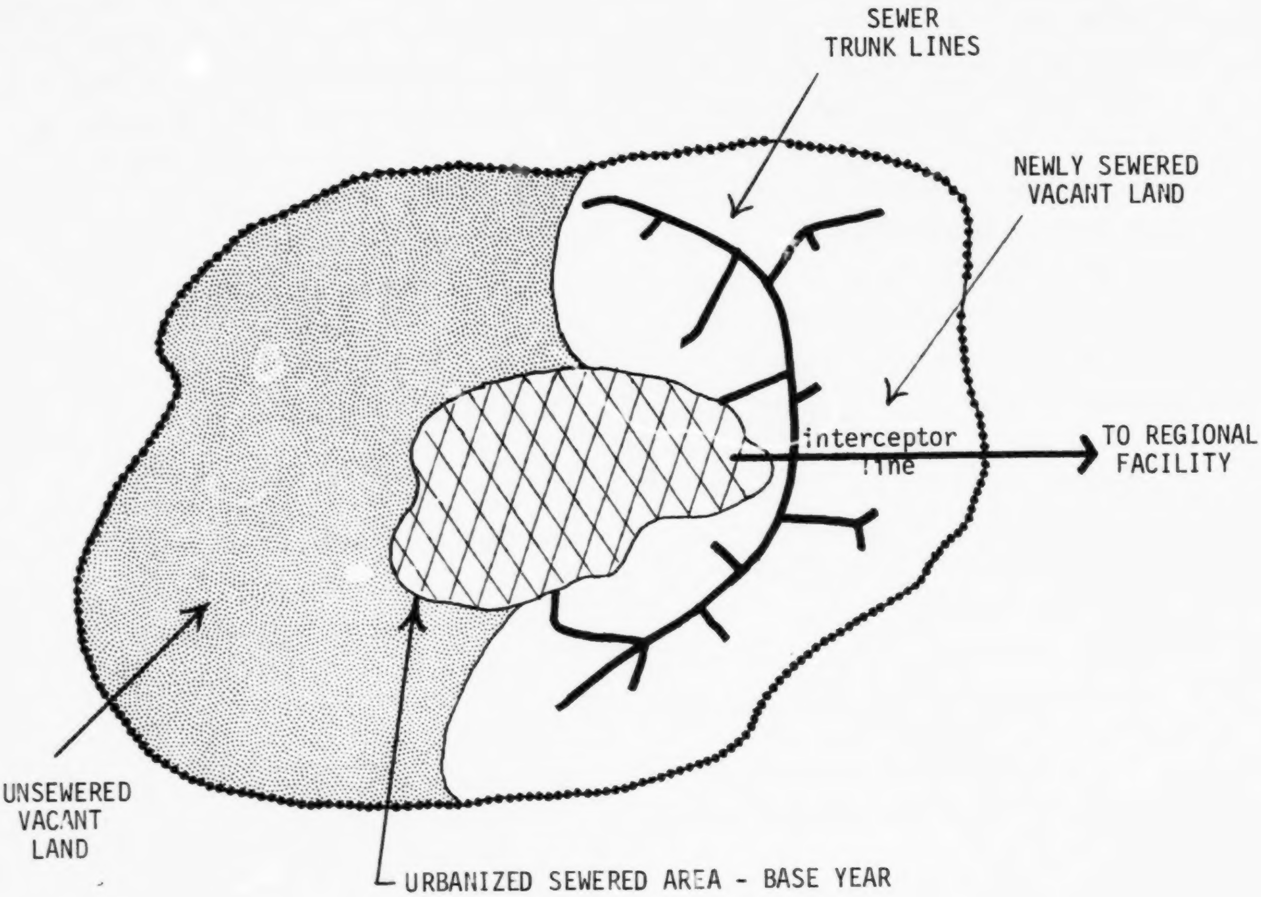

SEWER
TRUNK LINES
NEWLY SEWERED
VACANT LAND
interceptor
line
TO REGIONAL
FACILITY
UNSEWERED
VACANT
LAND
URBANIZED SEWERED AREA - BASE YEAR

FIGURE 2-1
EXAMPLE OF THE DISPLACED GROWTH PHENOMENON WITHIN A HYPOTHETICAL DRAINAGE BASIN

effects are: (1) those to simply upgrade treatment technology (e.g., from primary to secondary, or from direct discharge of raw sewage to any level of treatment), or (2) those that increase plant capacity only to more nearly match collection network capacity, e.g., to handle periods of high flow during wet weather from combined storm-sanitary sewers (preventing overflow conditions). The principal criteria which describe the definition of major projects can be categorized as type, size, and timing requirements and are discussed separately below.

Wastewater projects can consist of either new construction of, or extensions to: (1) treatment plants, (2) interceptor lines and collection networks, (3) outfall sewers (discharge lines from the treatment plant, if it exists, to the receiving body of water), or (4) waste stabilization ponds and appurtenances. We are not interested in the last two project types, since these involve the treatment and discharge of wastes, and do not affect the availability of sewer connections. Project type (1), expansion of treatment plant capacity, is also, *by itself*, not an element that will likely increase land development. When collection network capacity exceeds treatment capacity, rarely are additional sewer connections prohibited [60]. And if restrictions are enacted, such moratoria are generally difficult to maintain, even in cases of treatment plant overloading. Courts are unlikely to allow such moratoria to stand more than a year or two since sewers are considered a public asset [64]. Thus, collection network capacity is a greater growth constraint than treatment plant capacity. Consequently, an increase in the treatment capacity of a wastewater project generally does not increase its growth-inducing effects, unless the collection network is concurrently or subsequently expanded as well. Project type (2), expansion of the collection network, is the wastewater project type we are primarily interested in for the purposes of defining a major project. Past research has shown that the major secondary impacts associated with wastewater facilities "result from the placement, sizing and staging of interceptor sewers, and the provision of reserve capacity in these sewers" [98]. Usually, a project involves the construction or extension of various combinations of the four project *types*. Thus, although our principal interest is in the expansion of wastewater collection systems, changes in treatment and disposal of wastes may also be involved.

A minimum *size* requirement was necessary to ensure a wastewater project is "major", i.e., large enough to have had the potential, upon construction, for inducing a "significant quantity" of land development in its legal service area. Based upon discussions with EPA personnel from the Municipal Construction Division and the Office of Transportation and Land Use Policy, the conclusion was reached that wastewater projects with a collection network capacity below 1 MGD (million gallons per day) are too small to induce a significant quantity of land development. Thus, this figure was chosen as one minimum size requirement. Since sewage flow typicably varies between 90 and 200 gallons per capita per day, the upper limit is generally used in wastewater system design [99,100]. Thus, a capacity of 1 MGD will fulfill the wastewater needs of approximately 5,000 persons. A second size requirement was that the wastewater project had to cost at least $200,000. Interceptor sewer projects for the period 1960-1970 cost typically in the range of $100,000 to $10,000,000 to construct [101]. The objective of the minimum cost requirement was again to eliminate small projects which could not include a detectable amount of land development.

*Timing* requirements were set to ensure that project impacts had been realized by 1970, the year in which final land use in the area of influence was measured. These requirements were specified as a 2-year time period centered around the base year of 1960 during which construction on the major project had to be essentially completed, i.e., from 1958 to 1962. Although choosing a wider time period would increase the availability of case study major projects, it would also increase the unexplainable variance in the model. This criteria was left open to revision, subject to the availability of sufficient case study projects which could satisfy it. Since construction completion dates may not always be available from information sources used to select case study major projects, an alternative timing requirement was set. The project had to receive matching federal grants (if allocated) between 1956 to 1960. The offset of 2 years from the construction period requirement of 1958 to 1962 is to account for two effects. First, there is usually a few months delay between the time a grant is made and construction begins (although this lag can vary from 1 month to 2 years). Second, the average construction period for a project costing over $200,000 is 1-2 years, see Table 2-2 (although it can range anywhere from 10 months to 3 1/2 years).

## TABLE 2-2

## POINT ESTIMATES OF THE CONSTRUCTION PERIOD OF INTERCEPTOR SEWER PROJECTS

| Total Project Cost 1967-69 $ | Months to Complete Construction Lower Limit | Average | Upper Limit |
|---|---|---|---|
| $ 25,000 | 4.7 | 7.1 | 10.8 |
| 50,000 | 6.1 | 9.3 | 14.1 |
| 100,000 | 7.7 | 11.7 | 17.9 |
| 250,000 | 10.1 | 15.4 | 23.5 |
| 500,000 | 12.1 | 18.5 | 28.1 |
| 1,000,000 | 14.3 | 21.8 | 33.1 |
| 2,000,000 | 16.5 | 25.2 | 38.3 |
| 3,000,000 | 17.9 | 27.3 | 41.5 |

Source: [101]

Another timing problem which was encountered is related to the fact that wastewater construction projects are often constructed in several phases, several years apart. This project *phasing* is quite common and is caused by: (1) the large costs of such projects (despite federal matching funds), (2) the reluctance of communities to finance a large debt all at once, and (3) projected demand over time. For example, the City of Denton, Texas (1970 population of 40,000) completed four different wastewater projects within the period of 1960-1970 [52]. In 1960, construction began on a $1.2 million project to extend interceptor sewer lines and construct a new city treatment plant. In 1963, the collection network was again expanded at a cost of $230,000. In 1966, a $260,000 project constructed additional sewer outfalls and in 1968 a $1.3 million addition was made to the sewage treatment plant. For our purposes, only the first two projects are of interest since they alone affected the availability of sewer connections. To eliminate the 1960 project in Denton from consideration as a case study because of phasing would be a mistake, however, since initiation of the second project could have been due to extensive development pressures in the community (population increased 49% from 1960 to 1970). Thus, communities in which most of the development potential of a wastewater major project is rapidly used may experience project phasing more often than other communities. To exclude such phasing might bias the resultant model. Therefore, we extended the definition of wastewater major project to allow phased projects on a limited basis, by considering projects in which a second phase of collection network construction was completed by 1965. In such cases, the system capacity was defined as that present after the second phase of construction was completed. This approach treats the construction as really one planned project, spread over time. The definition also allows 5 years for the induced growth effects associated with the second phase to be realized.

In summary, the definition determined was:

> A wastewater major project is considered to involve principally the construction or extension of interceptor or collector sewer lines during the period 1958 to 1962 in a community in the United States. If construction date information is not available, then a grant funding date in the period 1956-1960 is acceptable. The project had to affect an increase in absolute system collection capacity of 1 MGD or more, and had to cost a minimum of $200,000 to construct. Phased projects are considered if the first phase of construction on the collection network was complete within 1958-1962 and the last phase of construction was complete by 1965.

## F. DEFINE AREA OF ANALYSIS

Careful study was made of the size and characteristics desired for the area of analysis to be used in this study to measure the induced land use impacts of major wastewater projects. In the GEMLUP-I project [17], a fixed area of 10,000 acres was used because it "captured" most secondary development (75-80%) associated with the project types investigated. In the current study, however, the "size" of a wastewater major project relates to its legal service area, which can vary from few hundred to several hundred thousand acres (see Table 2-3). Thus, it was necessary to use a variable area of analysis to accommodate such wide variability in project size.

For a given major project, selecting the size of the area of analysis required cognizance of two potential problems. First, if the area selected was too large, the influence of the major project on the end state land use would be small compared to the influence of regional growth factors. Thus, although the model might make accurate predictions, it would not be simulating principally the induced growth effects of the major project. On the other hand, if the area selected was too small, the model might not represent all of the associated induced development and again it would be underestimating the major project's influence. The approach employed in defining the area of analysis involved determining first the area in which most secondary land use occurs, and second, the area in which most tertiary land use, induced by secondary development, occurs and whether this extends beyond the boundaries of the secondary development. Note that the partitioning of induced

## TABLE 2-3

## CONSTRUCTION GRANT INFORMATION ON A SAMPLE OF CURRENT WASTEWATER PROJECTS IN FEDERAL REGIONS II, IV AND VI

| | | Service Area (acres) | Percent Developable | Initial Population | Design Population |
|---|---|---|---|---|---|
| Number of Observations | | 49 | 50 | 51 | 56 |
| Mean | | 20,118 | 60 | 26,894 | 65,459 |
| Standard Deviation | | 33,267 | 22 | 40,764 | 91,470 |
| Range | | 200 to 149,000 | 0 to 96 | 0 to 173,500 | 1,470 to 402,000 |
| Distribution | 18% | <1,500 | | | |
| | 18% | 1,500-3,000 | | | |
| | 23% | 3,000-10,000 | | | |
| | 25% | 10,000-50,000 | | | |
| | 8% | 50,000-100,000 | | | |
| | 10% | >100,000 | | | |

Source: [102].

development into secondary and tertiary types is done here only for the purpose of determining the areal extent of growth impacts from a wastewater major project. The implied temporal ordering is not represented explicitly in the GEMLUP model.

All secondary land use, induced directly by the wastewater major project, will occur within what is called the "legal service area" of that project. Since it is, in general, not possible to deny a developer access to a publicly-financed sewer with excess capacity [64], the "legal service area" is defined as the watershed boundaries of the area that drains, by gravity, to any point along the interceptor line, plus any areas outside the watershed connected to the collection network by pumping stations and force mains. Service of this entire area is not implausible since widely used sewer design standards require that interceptor pipes be sized to accommodate the full development of the drainage basin served [13]. In the case of essentially flat terrain (e.g., river deltas), the vertical inclination of pipes provides the necessary gravity flow for the system. Due to the costs of excavating, in these cases, the legal service area is usually restricted to an area no greater than 1,000 feet from any point along the interceptor line. The legal service area is, in general, not easy to map since watershed boundaries pursue a course independent of municipal boundaries. A single watershed may contain many towns, and a single town may overlay portions of several watersheds (e.g., the City of Lexington, KY has 7 different watersheds [104]). Also, any watershed can be partitioned into sub-drainage basins. Where an interceptor sewer only penetrates one end of a watershed, its legal service area may contain only a few of the sub-drainage areas in the watershed.

The legal service area defines the area in which secondary development is *possible*. There is a smaller area, contained within the legal service area, however, in which such development is most *probable*. An example of this is illustrated in Figure 2-2 where the legal service case is large and only the "skeleton" of a collection network is constructed. Specifically, the extent of the publicly constructed wastewater collection network is a single interceptor line placed parallel to the principal tributary (river) down the middle of the watershed, with only a few cross-connecting collector

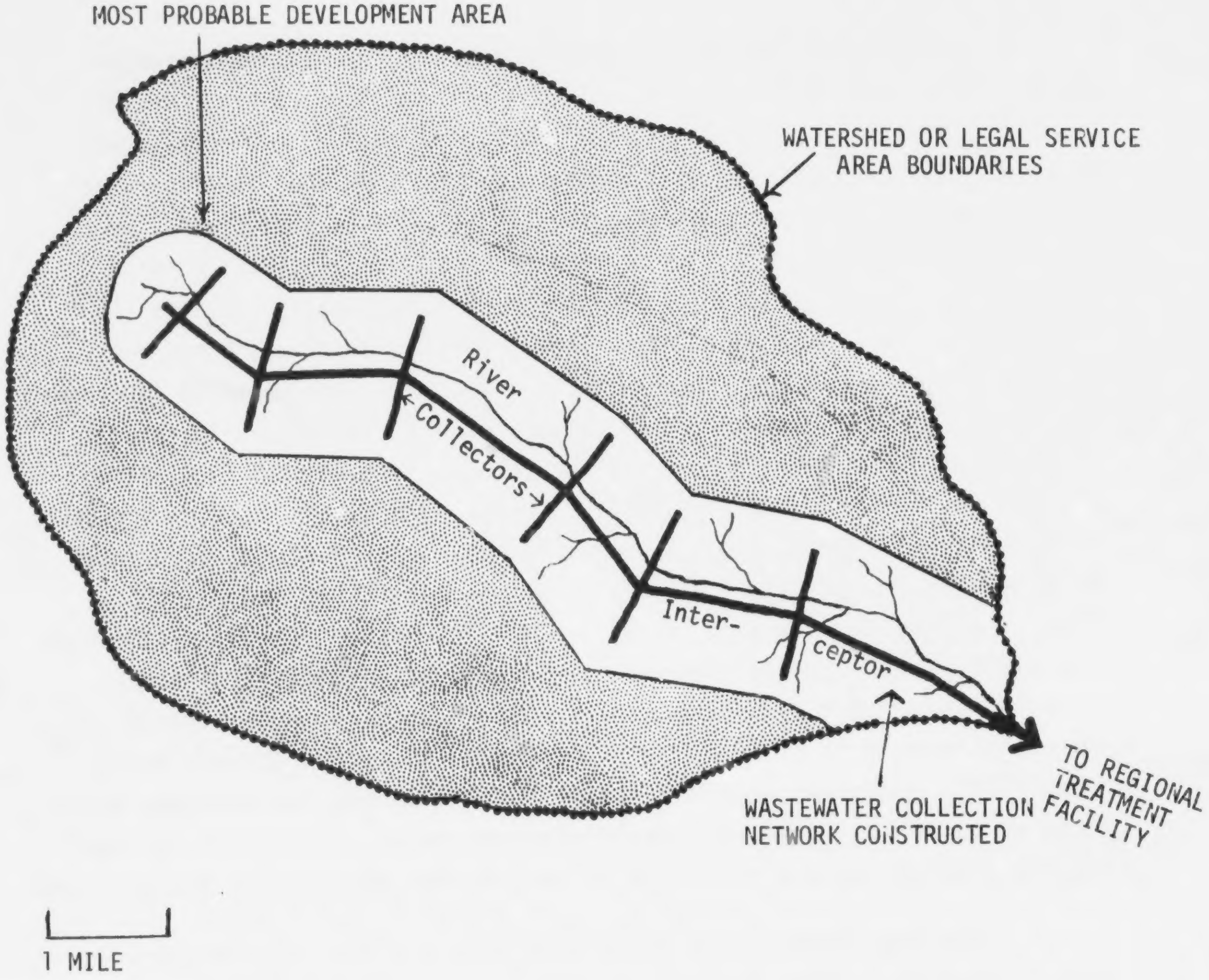

MOST PROBABLE DEVELOPMENT AREA
WATERSHED OR LEGAL SERVICE AREA BOUNDARIES
River
Collectors
Inter-ceptor
TO REGIONAL TREATMENT FACILITY
WASTEWATER COLLECTION NETWORK CONSTRUCTED
1 MILE

FIGURE 2-2

HYPOTHETICAL LEGAL SERVICE AREA OF A WASTEWATER MAJOR PROJECT ILLUSTRATING THE LOCATIONAL EFFECTS OF THE PROJECT'S COLLECTION NETWORK lines. In this case, the area closest to the river (including its flood plain) has the maximum potential for development due to the easy access to the collector network. Conversely, the area farthest from the interceptor line, near the watershed boundary, has the lowest development potential due to the cost a developer must incur to construct a wastewater line down to the existing network. This example illustrates the idea that the quantity of induced development is some declining function of the distance from the wastewater collection network. Promise and Leiserson [53] have studied this effect through a statistical analysis of housing data for the period 1963 to 1970 in suburban Washington, D.C. (Montgomery County, MD). The analysis considered the relationship between residential development and the distances to the nearest interceptor line and showed that the majority of development occurred within 5,000 feet (or 1 mile) of the line, with none beyond 10,000 feet (or 2 miles) distance. Work related to other land use types (e.g., industrial) was not found in the literature, and so these results were assumed to be representative of all secondary development.

The task of determining the size of the area in which tertiary land use occurs can be compared to the "area of influence" investigation that was performed in the GEMLUP-I study [17]. The factor inducing development is presumably newly constructed residential housing or industrial plants. In the GEMLUP-I study, it was assumed that a radius of 2 miles from the center of a Residential or Industrial major project encompassed most induced development. The issue here, however, is more complex. Unlike the previous study, it is generally not possible to define a center for the secondary development (the cause of tertiary development) since this can occur anywhere in a project's legal service area, but will most likely occur within 1 mile of an interceptor. This effect spreads out tertiary development over a wider area, at most, encompassed by the locus of points within 3 miles of an interceptor sewer line. However, the presence of more readily available sewer connections in the secondary development area simultaneously reduces the aerial extent of tertiary development. This, in conjunction with the fact that most tertiary development locates near its secondary development, led to the judgment that the composite of the secondary and tertiary land use areas generally defines an "area of analysis" no larger than the wastewater legal service area. Also, an expansion of the analysis beyond the legal service area to directly analyze the contribution of secondary growth to all tertiary uses would require a thorough analysis of all growth patterns in the region. For such an effort, a full regional land use model would have been required, along with extensive data for perhaps an hour's travel around the service area [105]. This was judged to be beyond both the scope of the study and the state-of-the-art (particularly in joint travel-location forecasting).

The wide variance in area of analysis size for major projects, shown in Table 2-3, indicates that small and large areas are likely to be mixed in the selection of case study projects. This causes a potential problem since small areas are less likely to include as much tertiary development as large areas. The interactions among induced land use types come in two forms, of which only one is significant for this project. On a large scale, interactions are determined by joint locational economies. For example, commercial uses need markets -- either households or businesses, and households need both housing and employment opportunities. These are the important interactions for this study. On a more detailed, local level, however, disruptions in these interactions can occur. Parcels of land often develop as all residential or industrial because of zoning and localized scale economies. Land is also commonly sold in large units, as agricultural activity gives way to urban development [106]. This leads to a market with local heterogenities and, when combined with on-site environmental impacts (particularly from industry and highways), to diseconomies of joint location on a local scale [61]. These two forces -- small scale homogeneity and large scale diversity -- confound each other in the analysis of small sites, and can lead to indeterminate effects. Since the GEMLUP model assumes a single causal hypothesis, a lower limit on area of analysis size of 5,000 acres was set to avoid such inconsistencies. This lower limit was based on the professional judgment of the staff after review of several wastewater major projects [102].

Another criteria related to the area of analysis involves the potential for induced development from a major project. Areas which are extensively urbanized in the base year will contain practically no vacant, developable land. In this case the relative growth effect of a major project will be imperceptible. Thus, the decision was made to exclude such cases by requiring the area of analysis to include significant amounts of vacant developable land. Also, some of this land must serve as a control area for measuring the displaced growth effects of the major project and so must be located more than 5,000 feet from the nearest interceptor sewer line.

Therefore, the following definition was used in this study:

> The area of analysis is defined as the legal service area of a wastewater major project in the base year. It must be a minimum size of 5,000 acres (or approximately 8 square miles), and contain significant amounts of vacant developable land, some of which must be more than 5,000 feet from the nearest interceptor line in the base year.

## G. INITIAL CHOICE OF VARIABLES

The objectives of this task were to: (1) define the initial list of endogenous land use variables in terms of standard Land Use Codes (LUC) [109], and (2) choose an initial set of exogenous model variables. Defining the endogenous variables in terms of an established classification system required a systematic review of each LUC category and an assessment of its suitability as a part of each GEMLUP land use variable. The results of this analysis are given in Table 2-4. Note that the "Office" variable name has been appended to reflect the existence of "Professional Services" in this variable, and that hospital land use has been moved to the "Office-Professional Services" variable. Since hospital land use was initially placed in a miscellaneous category, it was judged more appropriate to include it in a functionally related endogenous variable. Finally, as discussed in the section "Define Induced Development", the endogenous variables exclude the following land uses:

- LUC 4: Transportation, communication, and utilities (except LUC 45: Highway and street right-of-way)
- LUC 67: Governmental Services
- LUC 682 and 683: Universities and Trade schools
- LUC 72: Public assembly
- LUC 76: Parks
- LUC 8 and 9: Resource production and undeveloped land/water areas.

## TABLE 2-4

## DEFINITION OF ENDOGENOUS MODEL VARIABLES

| Name | Description | Land Use Codes [106] | |
|---|---|---|---|
| RES | Number of dwelling units per 10,000 acres of area of analysis in 1970. | 11 | Household units |
| | | 12 | Group quarters |
| | | 13 | Residential hotels |
| | | 14 | Mobile home parks or courts |
| | | 19 | Other residential |
| COMM | Commercial land use per 10,000 acres of area of analysis in 1970 in 1,000 square feet. | 52-59 | Retail trade |
| | | 62 | Personal services |
| | | 64 | Repair services |
| | | 66 | Contract construction services |
| OFFICE | Office-Professional services land use per 10,000 acres of area of analysis in 1970 in 1,000 square feet. | 61 | Finance, insurance, and real estate services |
| | | 631-636<br>638,639 | Business services (excludes warehousing) |
| | | 65 | Professional services |
| | | 692 | Welfare and Charitable services |
| | | 699 | Other services |
| MANF | Manufacturing land use per 10,000 acres of area of analysis in 1970 in 1,000 sq. ft. | 2,3 | Manufacturing |
| WHOLE | Wholesale-warehousing land use per 10,000 acres of area of analysis in 1970 in 10,000 square feet. | 51 | Wholesale trade |
| | | 637 | Warehousing and storage services |
| HIWAYS | Non-expressway highway lane miles per 10,000 acres of area of analysis in 1970. | 45 | Highway and street right-of-way |
| EDUC | Educational land use per 10,000 acres of area of analysis in 1970 in 1,000 square feet. | 681 | Nursery, primary, and secondary education |

## TABLE 2-4 (CONTINUED)
## DEFINITION OF ENDOGENOUS MODEL VARIABLES

| Name | Description | Land Use Codes [106] | |
|---|---|---|---|
| REC | Active, outdoor recreational land use per 10,000 acres of area of analysis in 1970 in acres. | 73 | Amusements |
| | | 74 | Recreational activities |
| | | 75 | Resorts and group camps |
| OTHER | Other urban land uses per 10,000 acres of area of analysis in 1970 in 1,000 square feet. | 15 | Transient lodgings |
| | | 691 | Religions activities |
| | | 71 | Cultural activities and nature exhibitions |
| | | 79 | Other cultural, entertainment, and recreational |

Next, a list of exogenous variables thought to be causally related to the endogenous land use variables was prepared. The types of exogenous variables that were required can be generally categorized as representing:

- Service area characteristics in the base year (1960)
- Regional growth factors for the time period of analysis (1960-1970)
- The development potential of the wastewater major project (1960).

The base year characteristics of the wastewater service area include measures of socioeconomic variables, accessibility, and land use constraints, and serve as surrogates for detailed estimates of base year land use, which could not be entered explicitly in the model due to the unavailability of data. Land use growth in the area of analysis is due to both regional activities and the development potential of the major project. Thus, exogenous variable categories have been specified for these major effects. The identification of exogenous variables in each of these categories was based on literature review and previous experience (i.e., professional judgement) of the staff in secondary impacts of infrastructure investments *. The exogenous variables selected for each of the three general categories are given in Table 2-5. Due to the wide variability in case study project area size, all exogenous variables, with the exception of AREA, are normalized to a percent or per acre basis. AREA has been included in the list of exogenous model variables to control for any variation in the general structure of causal relationships related to the variation in area of analysis scale.

The time period for the majority of exogenous variables is the year (t+0) when the wastewater major project was completed and became operational. Due to data constraints, however, published data sources (e.g., the U.S. Census) had to be used as the basis for several exogenous variables. In

---

*See references [17, 19, 53, 57-61, 63-66, 80, 107].

## TABLE 2-5

## DEFINITION OF EXOGENOUS MODEL VARIABLES

| Name - Description | Data Source |
|---|---|
| Service Area Base Year Characteristics: Socioeconomic Variables | |
| DUACRE = Dwelling units per mile$^2$ in area of analysis in 1960. | |
| = (100*du60)/acre | |
| where: du60 = 1960 census tracts* housing units in 100s | Census |
| acre = 1960 area of census tracts in miles$^2$ | Census |
| VACHSE = Percent vacant available dwelling units in area of analysis in 1960. | Census |
| INCOME = Relative medium income of families and unrelated individuals in area of analysis compared to county income levels in 1960. | |
| = (10*inc)/median | |
| where: inc = 1960 median income for families in census tracts in \$10s | Census |
| median = 1960 median income for families in the county** | Census |
| VACOFF = Vacancy rate of office buildings in area of analysis in 1960. | BOMA/Planning Agency |
| UNEMP = Unemployment rate in area of analysis (census tracts) in 1960. | Census |
| OFFJOB = Office employment per mile$^2$ in area of analysis in 1960. | |
| = (100*smoff)/acre | |
| where: smoff = 1960 office employment in census tracts in 100s***. | Census |
| WWJOB = Warehouse and wholesale employment per mile$^2$ in area of analysis in 1960. | |
| = (100*wwemp)/acre | |
| where: wwemp = 1960 employment in wholesale trade$^+$ in census tracts in 100s. | Census |

---

* Census tracts refers to those tracts most closely approximating the area of analysis in areal extent.

** County refers to the county containing most of the legal service area.

*** FIRE, Business Services, Public Administration, and Repair Services.

+ Trucking, Warehousing, and Wholesale Trade.

## TABLE 2-5 (CONTINUED)

## DEFINITION OF EXOGENOUS MODEL VARIABLES

| Name - Description | Data Source |
|---|---|
| MANJOB = Manufacturing employment per $\text{mile}^2$ in area of analysis in 1960.<br>= (100*manemp)/acre<br>where: manemp = 1960 manufacturing employment in census tracts in 100s. | Census |
| JOBS = Total employment per $\text{mile}^2$ in area of analysis in 1960.<br>(100*totemp)/acre<br>where: totemp = 1960 total civilian employment in census tracts in 100s. | Census |
| NONHH = Non-household population per $\text{mile}^2$ in area of analysis in 1960.<br>= (100*nonh60)/acre<br>where: nonh60 = 1960 population in group quarters in census tracts in 100s. | Census |
| POOR = Percent of total families with income below $3,000 in area of analysis (census tracts) in 1960. | Census |
| RENTS* = Percent of total housing units that are renter occupied in area of analysis in 1960. | Census |
| VALUE = Median value of housing units in area of analysis (census tracts) in 1960. | Census |
| ROOMS = Median number of rooms in housing units in area of analysis (census tracts) in 1960. | Census |
| KIDS = School age children per 100 households in area of analysis in 1960.<br>= 100*sch60/du60<br>where: sch60 = 1960 population 0-14 years of age in census tracts in 100s. | Census |

---

* Used only to predict the disaggregation percentage variables of SFDET, MF, and SFATT.

TABLE 2-5 (CONTINUED)

DEFINITION OF EXOGENOUS MODEL VARIABLES

| Name - Description | Data Source |
|---|---|
| UNIV = Categorical variable to indicate the presence or absence of a college or university in the area of analysis in 1960.<br>where: 1 = a college or university existed in census tracts in 1960.<br>0 = none existed. | Planning Agency |
| RELJOB = Relative employment density of the area of analysis in 1960.<br>= JOBS/(100*rempl/areal)<br>where: rempl = 1960 total employment in the SMSA in 100s.<br>areal = 1960 SMSA land area in $mile^2$. | Census<br>Census |
| RELWW = Relative warehouse and wholesale employment density of the area of analysis in 1960.<br>= WWJOB/(100*rwwemp/areal)<br>where: rwwemp = 1960 employment in warehouse and wholesale trade in the SMSA in 100s. | Census |
| RELOFF = Relative office employment density in the area of analysis in 1960.<br>= OFFJOB/(100*roffl/areal)<br>where: roffl = 1960 SMSA office employment in 100s. | Census |
| RELMAN = Relative manufacturing employment density of the area of analysis in 1960.<br>= MANJOB/(100*rempl*manper/areal)<br>where: manper = percent of 1960 total SMSA employment in manufacturing. | Census |
| GOVT = Total county government expenditures in 1962 in $10^6$ \$. | Census |
| LAND = Price of vacant land in area of analysis in 1960 relative to median regional income.<br>= price/median<br>where: price = median price of one acre of residential vacant land (\$) in area of analysis in 1960. | Planning Agency |

TABLE 2-5 (CONTINUED)

DEFINITION OF EXOGENOUS MODEL VARIABLES

| Name - Description | Data Source |
|---|---|
| POPDEN = Population density of the area of analysis in 1960 in persons per mile$^2$. | |
| = (100*pop)/acre | |
| where: pop = total population of census tracts in 1960 in 100s. | Census |
| SFDET = Percent of housing that is single family detached in area of analysis in 1970. | |
| = unit1/hse70 | |
| where: unit1 = number of single family detached units in census tracts in 1970 in 100s. | Census |
| hse70 = total number of housing units in census tracts in 1970 in 100s. | Census |
| MF = Percent of housing that is multifamily in area of analysis in 1970. | |
| = (unit34+unit5+unit50)/hse70 | |
| where: unit34 = number of 3 and 4 family housing units in census tracts in 1970 in 100s. | Census |
| unit5 = number of 5-49 family housing units in census tracts in 1970 in 100s. | Census |
| unit50 = number of 50$^+$ family housing units in census tracts in 1970 in 100s. | |
| SFATT = ((hse70-unit1)/hse70)-MF | |
| = Percent of housing that is single family attached in area of analysis in 1970. | |
| PCOMM1 = Percent of commercial development less than 50,000 ft$^2$ in floor area. | |
| = comm1/(comm1 + comm2 + comm3) | |
| where: comm1 = total commercial floor space in 1,000 ft$^2$ in area of analysis in 1970 for buildings with less than 50,000 ft$^2$. | Planning Agency |
| comm2 = total commercial floor space in 1,000 ft$^2$ in 1970 in area of analysis for buildings with less than 100,000 ft$^2$ but greater than 50,000 ft$^2$. | |
| comm3 = total commercial floor space in 1,000 ft$^2$ in 1970 in area of analysis for buildings with greater than 100,000 ft$^2$. | |

## TABLE 2-5 (CONTINUED)
## DEFINITION OF EXOGENOUS MODEL VARIABLES

| Name - Description | Data Source |
|---|---|
| PCOMM2 = Percent of commercial development with floor area between 50,000 and 100,000 $ft^2$.<br>= comm2/(comm1 + comm2 + comm3) | |
| PCOMM3 = Percent of commercial development with floor area greater than 100,000 $ft^2$.<br>= comm3/(comm1 + comm2 + comm3) | |
| Service Area Base Year Characteristics: Accessibility | |
| DISCBD = Distance in miles from centroid of area of analysis to centroid of nearest CBD* in year (t + 0)** | USGS topographic map |
| RRMILE = Railroad mileage per $mile^2$ in the area of analysis analysis in year (t + 0)<br>= (rail*640)/AREA<br>where: rail = railroad mileage in area of analysis in year (t + 0) | USGS topographic map |
| ACCESS = Limited-access highway interchanges per $mile^2$ in area of analysis in year (t + 5)<br>= (intchg*640)/AREA<br>where: intchg = number of limited access interchanges in area of analysis in year (t + 5) | USGS topographic map/ Planning Agency |
| INTDEN = Relative limited-access highway interchange density of the area of analysis in year (t + 5)<br>= ACCESS/(ctyacc*640/county)<br>where: ctyacc = number of limited access interchanges in the county in year (t + 5)<br>county = area of county in $mile^2$ | USGS topographic map/ Planning Agency |
| TRANS = Number of transit stops (bus and commuter rail) in the area of analysis in year (t + 0) | Planning Agency |
| AIRPRT = Distance in miles from centroid of area of analysis to centroid of nearest commercial airport in the year (t + 0) | USGS topographic map |

---

* Central Business District, defined as the center of the nearest urban area with population exceeding 100,000.

**t is the year the wastewater major project was completed and became operational.

## TABLE 2-5 (CONTINUED)

## DEFINITION OF EXOGENOUS MODEL VARIABLES

| Name - Description | Data Source |
|---|---|
| Service Area Base Year Characteristics: Land Use Constraints | |
| AREA = Area of analysis in acres | Project Data |
| VACANT = Percent vacant developable acreage in area of analysis in year (t + 0)<br>= vacdev/(AREA-vacund) | |
| where: vacdev = vacant developable acreage in area of analysis in year (t + 0) | Planning Agency |
| vacund = vacant undevelopable acreage in area of analysis in year (t + 0) | Planning Agency |
| RZONED = Percent of total acreage zoned for residential use in the area of analysis in year (t + 0)<br>= rzone/AREA | |
| where: rzone = acres of land zoned for residential use in the area of analysis in year (t + 0) | Planning Agency |
| CZONED = Percent of total acreage zoned for commercial use in the area of analysis in year (t + 0)<br>= czone/AREA | |
| where: czone = acres of land zoned for commercial use in the area of analysis in year (t + 0) | Planning Agency |
| OZONED = Percent of total acreage zoned for office use in the area of analysis in year (t + 0)<br>= ozone/AREA | |
| where: ozone = acres of land zoned for office use in the area of analysis in year (t + 0) | Planning Agency |
| IZONED = Percent of total acreage zoned for industrial use in the area of analysis in year (t + 0)<br>= izone/AREA | |
| where: izone = acres of land zoned for industrial use in the area of analysis in year (t + 0) | Planning Agency |
| SOILS = Percentage of total area of analysis having a "severe" soil type classification for urban development suitability<br>= soil/AREA<br>where: soil = Acreage in area of analysis with a "severe" soil type classification with regard to suitability for urban development. | Planning Agency/Soil Conservation Service (SCS) |

TABLE 2-5 (CONTINUED)

DEFINITION OF EXOGENOUS MODEL VARIABLES

| Name - Description | Data Source |
|---|---|
| ONLOT = Percentage of total area of analysis having a "severe" soil type classification for on-lot sewage disposal suitability.<br>= onsite/AREA<br>where: onsite = Acreage in area of analysis with a "severe" soil type classification for on-lot sewage disposal suitability. | Planning Agency/SCS |
| LIMITS = Categorical variable to indicate the severity of governmental restrictions on on-lot sewage disposal during the period (t + 0) to 1970.<br>where: 4 = on-lot disposal is prohibited entirely.<br>3 = on-lot disposal is prohibited except on large lots.<br>2 = on-lot disposal is permitted but percolation tests are required.<br>1 = on-lot disposal is permitted but package plants are prohibited.<br>0 = no restrictions. | Planning Agency/ Local Government |
| POLICY = Categorical variable to indicate the presence or absence of governmental policies designed to limit the number of hookups to the sewerage system in area of analysis anytime during the period (t + 0) to 1970. Examples are sewer moratoria and rationed connections.<br>where: 1 = policies on hookup limitations existed<br>0 = policies did not exist. | Planning Agency/ Local Government |
| TLIMIT = The number of years during the period of analysis that on-site sewage disposal was limited<br>= 1970-ylimit<br>where: ylimit = the year on-site disposal limitations went into effect (or 1970 if no restrictions). | Planning Agency/ Local Government |
| TPOLIC = The number of years during the period of analysis that sewer system hookup limitations were in effect<br>= 1970-ypolic<br>where: ypolic = the year limitations on sewage system hookups went into effect (or 1970 if no restrictions) | Planning Agency/ Local Government |

TABLE 2-5 (CONTINUED)

DEFINITION OF EXOGENOUS MODEL VARIABLES

| Name - Description | Data Source |
|---|---|
| Regional Growth Factors | |
| ENERGY = Relative electrical energy cost factor in municipality compared to average U.S. commercial rate in 1960. | Census |
| = encost/$51.59 | |
| where: encost = cost of energy for commercial users in 1960 in units of dollars per 1500 KWh. | |
| DRIVE = 100s of workers who drive per mile$^2$ in the county in 1960. | |
| = (autoco)/county | |
| where: autoco = workers who drive to work in county in 1960 in 100s. | Census |
| RIDET = Workers who ride mass transit per mile$^2$ in the county in 1960. | |
| = massco/county | |
| where: massco = workers who use mass transit in county in 1960 in 100s. | Census |
| PERCHG = Percent change in county population 1960-1970 | Census |
| DENCHG = Population density change in the county 1960-1970. | |
| = (copop2/county - copop1/county) | |
| where: copop2 = county population in 1970. | Census |
| copop1 = county population in 1970. | Census |
| JOBCHG = Change in total regional employment per mile$^2$ 1960-1970. | |
| = 100*(remp2 - remp1)/SMArea | |
| where: remp1 = 1960 total employment in SMSA in 100s. | Census |
| remp2 = 1970 total employment in SMSA in 100s. | Census |
| SMArea= area of SMSA in miles$^2$. | Census |
| HSECHG = Percent change in total regional housing units 1960-1970. | |
| = (rhse2 - rhse1)/rhse1 | |
| where: rhse1 = total housing units in SMSA in 1960 in 100s. | Census |
| rhse2 = total housing units in SMSA in 1970 in 100s. | Census |

## TABLE 2-5 (CONTINUED)

## DEFINITION OF EXOGENOUS MODEL VARIABLES

| Name - Description | Data Source |
|---|---|
| COMCHG = Percent change in total regional retail trade* employment 1960-1970. | |
| = (rcom2 - rcom1)/rcom1 | |
| where: rcom1 = total retail trade employment in SMSA in 1960 in 100s. | Census |
| rcom2 = total retail trade employment in SMSA in 1970 in 100s. | Census |
| POPDIF = Percent change in regional population 1960-1970. | Census |
| MANCHG = Percent change in regional manufacturing employment 1960-1970. | |
| = (perman*remp2-manper*remp1)/(manper*remp1) | |
| where: perman = % of 1970 total SMSA employment in manufacturing | Census |
| manper = % of 1960 total SMSA employment in manufacturing | Census |
| SERCHG = Percent change in regional services** employment 1960-1970. | |
| = (rser2 - rser1)/rser1 | |
| where: rser1 = 1960 SMSA service employment in 100s. | Census |
| rser2 = 1970 SMSA service employment in 100s. | Census |
| HOSCHG = Percent change in regional hospital employment 1960-1970. | |
| = (rhosp2 - rhosp1)/rhosp1 | |
| where: rhosp1 = 1960 SMSA hospital employment in 100s. | Census |
| rhosp2 = 1970 SMSA hospital employment in 100s. | Census |
| EDUCHG = Percent change in regional educational employment 1960-1970. | |
| = (red2 - red1)/red1 | |
| where: red1 = 1960 SMSA educational employment in 100s. | Census |
| red2 = 1970 SMSA educational employment in 100s. | Census |
| OFFCHG = Percent change in regional office*** employment 1960-1970. | |
| = (roff2 - roff1)/roff1 | |
| where: roff1 = 1960 SMSA office employment in 100s. | Census |
| roff2 = 1970 SMSA office employment in 100s. | Census |

---

* Food and Dairy, Eating and Drinking, and Other Retail.

** Professional and Related Services, Other Personal Services, Entertainment, and Welfare and Fraternal Organizations.

*** FIRE, (Finance, Insurance, and Real Estate) Business Services, Public Administration, and Repair Services.

TABLE 2-5 (CONTINUED)

DEFINITION OF EXOGENOUS MODEL VARIABLES

| Name - Description | Data Source |
|---|---|
| PERCAP = Median family income in SMSA in 1970. | Census |
| EMPOP = Regional (SMSA) employment to population ratio in 1970. | Census |
| = remp2/rpop2<br>where: rpop2 = 1970 SMSA population in 100s. | Census |
| CAPCHG = Change in median family SMSA income 1960-1970.<br>= PERCAP - per60<br>where: per60 = median family income in SMSA in 1960. | Census |
| STAY = Index of mobility - % of 1960 families who were in the same house in 1955. | Census |
| Wastewater Treatment Major Project Characteristics | |
| TIME = Number of years available for secondary growth to occur.<br>= phyear-year | |
| where: phyear = the year aerial photographs were taken from which land use data was extracted. | Planning Agency |
| year = the year construction was completed on the major project on initial phase. | Project Data |
| PHASE = Categorical variable to indicate the presence or absence of phasing of the major project.<br>where: 1 = phasing has occurred.<br>0 = phasing has not occurred. | Project Data |
| COST = Normalized per capita local costs ($) of the major project in area of analysis in year (t + 0).<br>= 1,000* (totcst/pi1 - fedcst/pi2)/popcom | |
| where: totcst = total major project construction cost (1,000 $) | Project Data |
| fedcst = federally funded share of major project cost (1,000 $) | Project Data |
| pi1 = consumer price index for "year" | |
| pi2 = consumer price index for the year of federally funding of the major project | |
| popcom = population served by major project facility in year (t + 0) | Project Data |
| LENGTH = Running length of interceptor sewer lines in miles going through relatively undeveloped land (<.1 du per acre) in area of analysis in year (t + 0). | Project Data |

TABLE 2-5 (CONTINUED)

DEFINITION OF EXOGENOUS MODEL VARIABLES

| Name - Description | Data Source |
|---|---|
| CROSS = Index of available undeveloped land in the area of analysis through which interceptor sewers go through in t + 0.<br>= 640*LENGTH/AREA | |
| SERVED = Percent of the area of analysis easily served by the major project in year (t + 0)<br>= area5k/AREA<br>where: area5k = acres of land within 5,000 feet of the major project interceptor sewer in area of analysis in year (t + 0) | Project Data |
| CAPAC1 = Total hydraulic design capacity of the major project wastewater collection system in million gallons per day (mgd) in year (t + 0) (or in 1965 if a phased project) | Project Data |
| CAPAC2 = Total hydraulic design capacity of the major project wastewater treatment plant in mgd in year (t + 0) | Project Data |
| PEAK = Actual Peak flow in the major project wastewater system in mgd in year (t + 0) | Project Data |
| RECAP = Reserve capacity of the wastewater major project collection system in year (t + 0)<br>= CAPAC1 - PEAK | |
| RECAP1 = Percent reserve capacity of the wastewater major project collection system in year (t + 0)<br>= 100* RECAP/PEAK | |
| RECAP2 = Percent reserve capacity of the wastewater major project treatment plant in year (t + 0)<br>= 100* (CAPAC2 - PEAK)/PEAK | |

these cases, the time period for which the variables defined is 1960, the average value of (t+0). In addition, the variable ACCESS is defined for the year (t+5) since the construction of new limited-access highway interchanges is commonly known five years in advance. Note that regional growth variables are based on standard Metropolitan Statistical Areas (SMSAs) which are sufficiently large so that measures of regional growth processes are essentially independent of the induced growth effects of major projects [102].

The variable PEAK required to define the exogenous variables RECAP, RECAP1 and RECAP2 may not always be available directly from project data files. In case it is not, the actual peak flow in gallons per day in the major project wastewater collection system in 1960 may be estimated by first computing the average flow as follows:

$$\begin{aligned}\text{Average flow} = {} & \Sigma\ (\text{industrial flows}) + \\ & 50 * (\text{1960 population in service area}) + \\ & 300 * \sum_i (\text{pipe length}_i * \text{pipe diameter}_i)\end{aligned}$$

The factor of 50 gallons per day is an average per capita use [108]. The last term computes infiltration flow, using a factor of 300 gallons per day [108] per inch diameter per mile of pipe, summing over all major sewer pipes. Peak flow is then obtained by applying the scaling factor given in Figure 2-6 of reference [100].

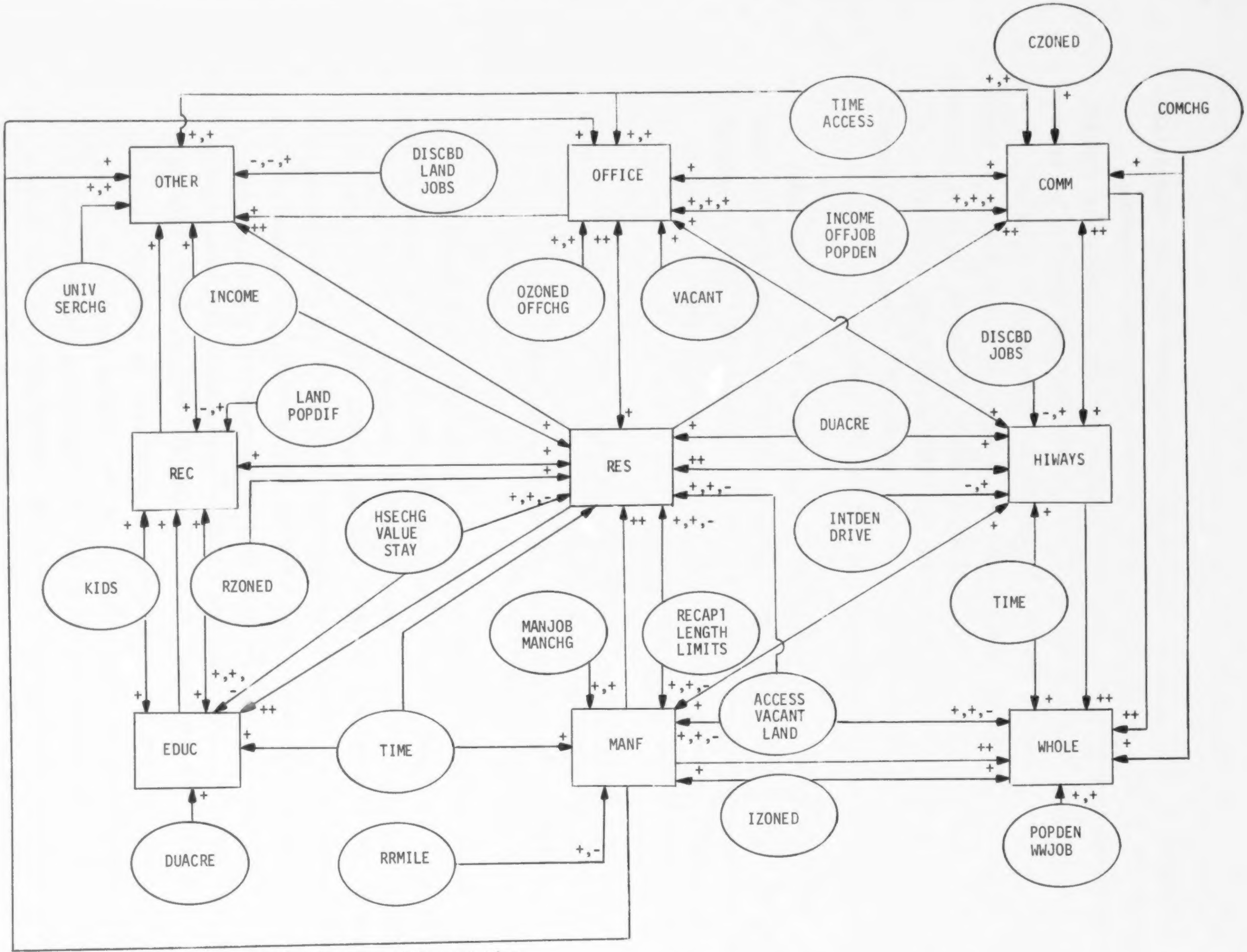
## H. SPECIFY INITIAL PATH ANALYTIC LAND USE MODEL

The objective of this task was to formalize the factor theory of induced development in a causal path diagram. Path diagrams are a schematic representation of the hypothesized causal relations among variables. Causality is shown by a single-headed arrow (or path) and simultaneity (reciprocal causation) between two variables is shown by a double-headed arrow. Use of this technique requires two assumptions about causality in the system: (1) a "weak" causal order exists among the variables and it is known, and (2) relationships among variables are causally closed. Weak causal ordering exists in a two variable set, $X_i$ and $X_j$, if it is known on logical, empirical, or theoretical grounds that $X_i$ (the exogenous variable) *may* affect $X_j$ (the endogenous variable) and that $X_j$ *cannot* affect $X_i$. Causal closure is simply the concept that, given a weak causal ordering ($X_i \rightarrow X_j$), the observed covariation between the two variables must be due only to the causal dependence of $X_j$ on $X_i$, their mutual dependence on some outside variable(s), or a combination of these factors [92]. Since path analysis is only capable of testing theories, rather than deducing them, the most comprehensive model possible was initially hypothesized. This ensures that the final path analytic model will represent the causal structure closest approximating the real world.

The approach to this task was to take the list of initial model variables previously compiled, and using our knowledge of the infrastructure relationships involved, to examine all possible causal relationships between each endogenous variable and all other exogenous and endogenous variables. For each variable pair, we hypothesized:

- If a relationship is plausible and if so, why,
- The direction of causal action, and
- Whether the hypothesized relationship is "weak" or "strong" (subjectively).

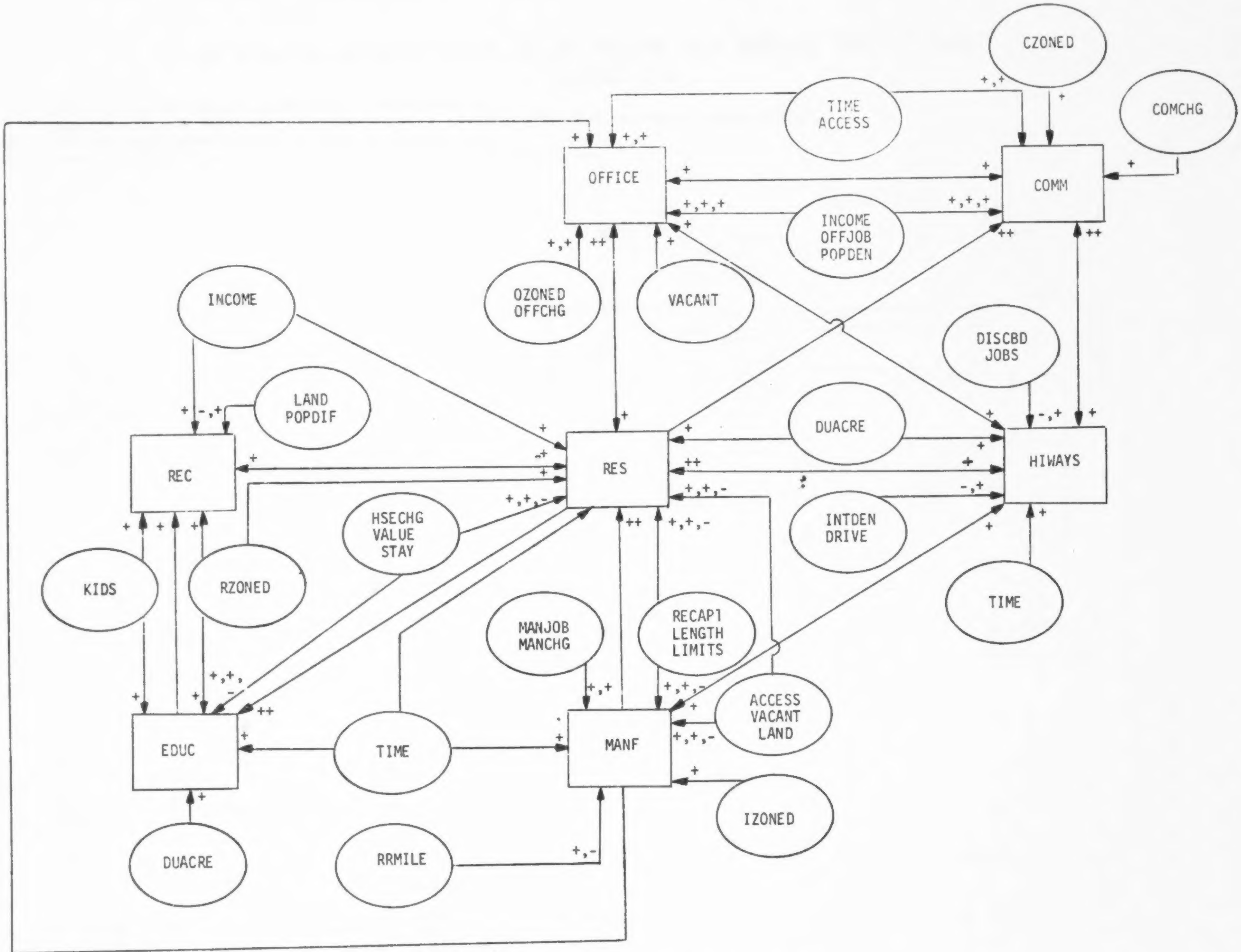
Where a causal relationship was perceived, it was coded into the path diagram shown in Figure 2-3, and the direction and strength of causal action are denoted by the "+" and "-" signs placed on the paths. Thus, Figure 2-3 represents the initial path analytic land use model. The specification of the model resulted in 9 structural equations of which 7 are simultaneous. These 7 equations, related to each other by interactions, are summarized in Tables 2-6 through 2-12 and the path diagram representing these equations is shown in Figure 2-4. The remaining 2 recursive equations are summarized in Tables 2-13 and 2-14 and Figure 2-5. The complete initial model contains 9 endogenous and 32 exogenous variables. An additional 34 exogenous variables serve as alternatives to those specified in the initial model (see Tables 2-6 through 2-14). The choice between prime and alternative exogenous variables was made based on the results of a zero-order correlation analysis for all model variables (see Section IV). This analysis also provided the necessary data to check for the problems of multicollinearity and suppressor variables.


CZONED
COMCHG
TIME ACCESS
OTHER
DISCBD LAND JOBS
OFFICE
COMM
INCOME OFFJOB POPDEN
UNIV SERCHG
INCOME
OZONED OFFCHG
VACANT
DISCBD JOBS
LAND POPDIF
REC
RES
DUACRE
HIWAYS
HSECHG VALUE STAY
INTDEN DRIVE
KIDS
RZONED
TIME
MANJOB MANCHG
RECAP1 LENGTH LIMITS
ACCESS VACANT LAND
EDUC
TIME
MANF
WHOLE
IZONED
DUACRE
RRMILE
POPDEN WWJOB

Note: Some Variables Appear Twice for Ease of Presentation.
Path Coefficients Separated by Commas Refer to Multiple Exogenous Variables, Respectively.
Double "+" Signs Denote Strong Causal Relationships.

FIGURE 2-3

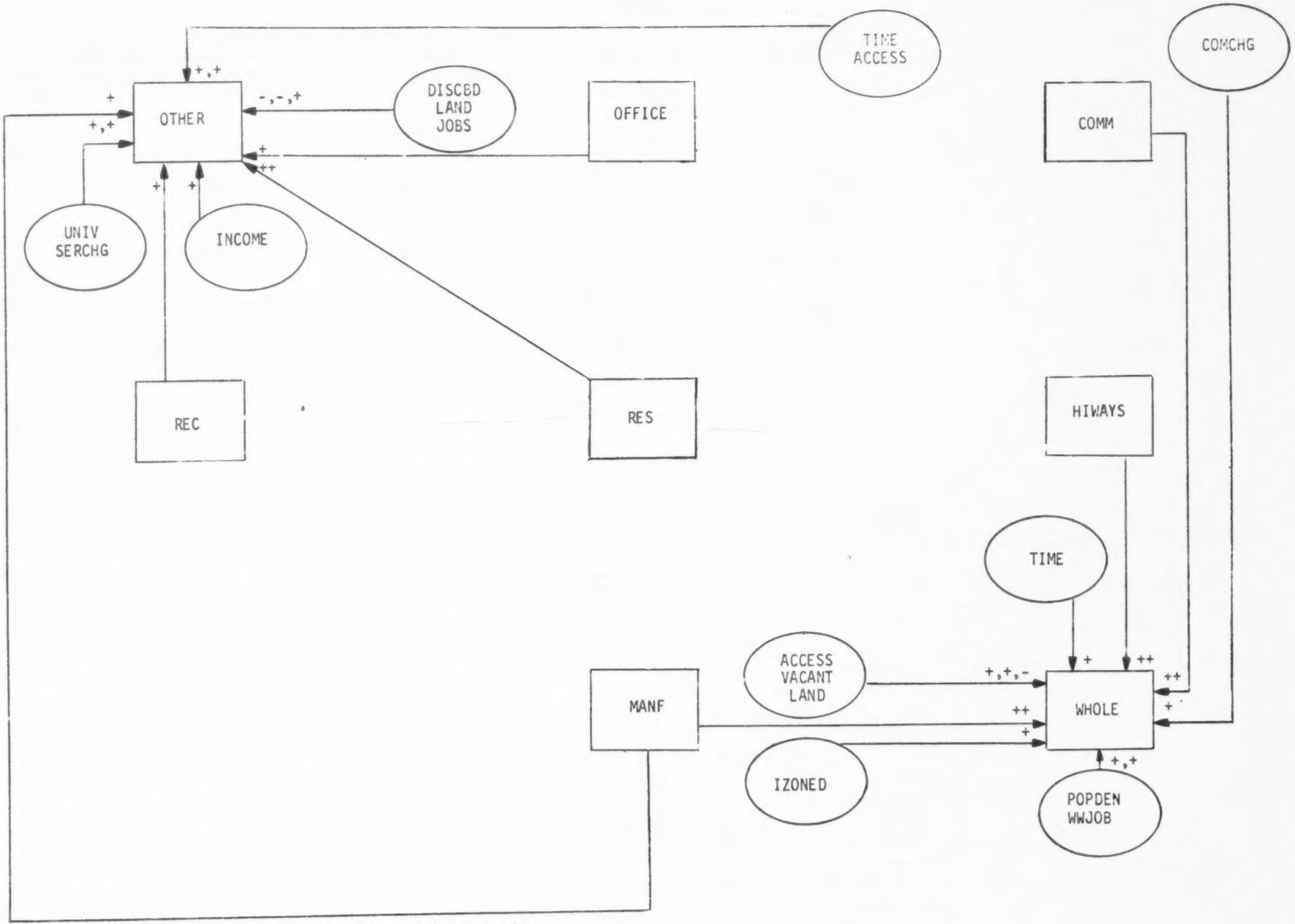
INITIAL SPECIFICATION OF THE PATH ANALYTIC LAND USE MODEL


CZONED
COMCHG
TIME ACCESS
OFFICE
COMM
INCOME OFFJOB POPDEN
OZONED OFFCHG
VACANT
INCOME
DISCBD JOBS
LAND POPDIF
DUACRE
REC
RES
HIWAYS
HSECHG VALUE STAY
INTDEN DRIVE
KIDS
RZONED
TIME
MANJOB MANCHG
RECAP1 LENGTH LIMITS
ACCESS VACANT LAND
EDUC
TIME
MANF
IZONED
DUACRE
RRMILE

Note: Some Variables Appear Twice for Ease of Presentation.
Path Coefficients Separated by Commas Refer to Multiple Exogenous Variables, Respectively.
Double "+" Signs Denote Strong Causal Relationships.

FIGURE 2-4

SEVEN EQUATION SIMULTANEOUS BLOCK IN THE INITIAL PATH ANALYTIC LAND USE MODEL


TIME ACCESS
COMCHG
+,+
+
+,+
OTHER
-,-,+
DISCBD LAND JOBS
OFFICE
COMM
+
++
+
+
UNIV SERCHG
INCOME
REC
RES
HIWAYS
TIME
ACCESS VACANT LAND
+,+,-
+
++
++
+
WHOLE
MANF
++
+
IZONED
+,+
POPDEN WWJOB

Note: Some Variables Appear Twice for Ease of Presentation.
Path Coefficients Separated by Commas Refer to Multiple Exogenous Variables, Respectively.
Double "+" Signs Denote Strong Causal Relationships.

FIGURE 2-5

TWO EQUATION RECURSIVE BLOCK IN THE INITIAL PATH ANALYTIC LAND USE MODEL

## TABLE 2-6

## INDEPENDENT VARIABLES FOR THE RES MODEL EQUATION

| Descriptor | Prime | Alternate(s) |
|---|---|---|
| Endogenous | +MANF | |
| Endogenous | +OFFICE | |
| Endogenous | +HIWAYS | |
| Endogenous | +REC | |
| MP* sewer capacity | +RECAP1 | +CAPAC1, +CAPAC2, +RECAP, +RECAP2, +PEAK |
| MP sewer service | + LENGTH | +CROSS, +SERVED, +COST |
| MP timing | +TIME | +PHASE |
| On-site disposal restrictions | -LIMITS | -POLICY, -TLIMIT, -TPOLIC |
| Population density | +DUACRE | +POPDEN |
| Regional influences | +HSECHG | +PERCHG, +POPDIF, +JOBCHG |
| Income | +INCOME | +PERCAP, +CAPCHG, -POOR, |
| Employment characteristics | +JOBS | +RELJOB, -UNEMP, -POOR |
| Housing characteristics | +VALUE | -VACHSE, +ROOMS, -NONHH |
| Land costs and taxes | -LAND | -GOVT |
| Mobility | -STAY | |
| Developable land | +VACANT | +AREA |
| Zoning | +RZONED | |
| Accessibility | +ACCESS | +INTDEN, +TRANS, +DRIVE, +RIDET |

* Major Project

## TABLE 2-7

## INDEPENDENT VARIABLES FOR THE COMM MODEL EQUATION

| Descriptor | Prime | Alternative(s) |
|---|---|---|
| Endogenous | +RES | |
| Endogenous | +HIWAYS | |
| Endogenous | +OFFICE | |
| Income | +INCOME | +PERCAP, +CAPCHG, -UNEMP, -POOR |
| Employment characteristics | +OFFJOB | +JOBS, RELOFF, +RELJOB |
| Population characteristics | +POPDEN | +UNIV, +NONHH |
| Accessibility | +ACCESS | +INTDEN, +TRANS, +DRIVE, +DISCBD, +RIDET |
| Zoning | +CZONED | +QZONED |
| MP timing | +TIME | |
| Regional influences | +COMCHG | +OFFCHG, +POPDIF |

TABLE 2-8

INDEPENDENT VARIABLES FOR THE
OFFICE MODEL EQUATION

| Descriptor | Prime | Alternative(s) |
|---|---|---|
| Endogenous | +COMM | |
| Endogenous | +RES | |
| Endogenous | +MANF | |
| Endogenous | +HIWAYS | |
| Income | +INCOME | +PERCAP, +CAPCHG |
| Employment characteristics | +OFFJOB | +RELOFF, +RELJOB, +JOBS |
| Population characteristics | +POPDEN | +NONHH |
| Accessibility | +ACCESS | +INTDEN, +TRANS, +DRIVE, -DISCBD, +RIDET |
| Zoning | +OZONED | +CZONED |
| MP timing | +TIME | |
| Regional influences | +OFFCHG | +JOBCHG, +POPDIF, HOSCHG |
| Development constraints | +VACANT | -VACOFF, -GOVT |

## TABLE 2-9

## INDEPENDENT VARIABLES FOR THE MANF MODEL EQUATION

| Descriptor | Prime | Alternative(s) |
|---|---|---|
| Endogenous | +HIWAYS | |
| MP sewer capacity | +RECAP1 | +CAPAC1, +CAPAC2, +RECAP2, +RECAP, +PEAK |
| MP sewer service | +LENGTH | +CROSS, +SERVED |
| Employment characteristics | +MANJOB | +JOBS, +RELJOB, RELMAN, +EMPOP, +UNEMP |
| Rail Accessibility | +RRMILE | |
| Road and Air Accessibility | +ACCESS | +INTDEN, -AIRPRT |
| Developable land | +VACANT | +AREA |
| Zoning | +IZONED | |
| On-site disposal restrictions | -LIMITS | -TLIMIT |
| MP timing | +TIME | |
| Land costs and taxes | -LAND | -GOVT |
| Regional influences | +MANCHG | +JOBCHG, +POPDIF |

## TABLE 2-10

## INDEPENDENT VARIABLES FOR THE HIWAYS MODEL EQUATION

| Descriptor | Prime | Alternative(s) |
|---|---|---|
| Endogenous | +RES | |
| Endogenous | +COMM | |
| Endogenous | +MANF | |
| Endogenous | +OFFICE | |
| Population characteristics | +DUACRE | +POPDEN, +POPDIF, +UNIV |
| Employment characteristics | +JOBS | +RELJOB, +MANCHG, +JOBCHG, +OFFCHG, +EMPOP |
| Motor vehicle density | +DRIVE | +RIDET |
| Distance to trip generators | -DISCBD | -AIRPRT |
| MP timing | +TIME | |
| Limited access highways | -INTDEN | -ACCESS |

## TABLE 2-11

## INDEPENDENT VARIABLES FOR THE EDUC MODEL EQUATION

| Descriptor | Prime | Alternative(s) |
|---|---|---|
| Endogenous | +RES | |
| Population characteristics | +DUACRE | -NONHH, +UNIV, +POPDEN |
| Housing characteristics | +VALUE | -VACHSE |
| School Age children | +KIDS | |
| Zoning | +RZONED | |
| MP timing | +TIME | |
| Mobility | -STAY | |
| Regional influences | +HSECHG | +PERCHG, +POPDIF, + EDUCHG |

## TABLE 2-12

## INDEPENDENT VARIABLES FOR THE REC MODEL EQUATION

| Descriptor | Prime | Alternative(s) |
|---|---|---|
| Endogenous | +RES | |
| Endogenous | +EDUC | |
| Income | +INCOME | +VALUE, -POOR, +GOVT |
| Population characteristics | +KIDS | -POPDEN, +UNIV |
| Land constraints | -LAND | +VACANT, +AREA |
| Zoning | +RZONED | |
| Regional influences | +POPDIF | +PERCHG |

## TABLE 2-13

## INDEPENDENT VARIABLES FOR THE WHOLE MODEL EQUATION

| Descriptor | Prime | Alternative(s) |
|---|---|---|
| Endogenous | +MANF | |
| Endogenous | +HIWAYS | |
| Endogenous | +COMM | |
| Employment characteristics | +WWJOB | +MANJOB, +RELJOB, RELWW, +RELMAN, +EMPOP |
| Land costs | -LAND | |
| Population characteristics | +POPDEN | |
| Accessibility | +ACCESS | +RRMILE, +INTDEN, -AIRPRT, -DISCBD |
| Developable land | +VACANT | +AREA |
| Zoning | +IZONED | |
| MP timing | +TIME | |
| Regional influences | +COMCHG | +POPDIF, +MANCHG |

## TABLE 2-14

## INDEPENDENT VARIABLES FOR THE OTHER MODEL EQUATION

| Descriptor | Prime | Alternative(s) |
|---|---|---|
| Endogenous | +RES | |
| Endogenous | +OFFICE | |
| Endogenous | +MANF | |
| Endogenous | +REC | |
| Income | +INCOME | +PERCAP, +CAPCHG, -POOR |
| Employment characteristics | +JOBS | +RELJOB, +EMPOP |
| Population characteristics | +UNIV | +POPDEN |
| Land constraints | -LAND | +VACANT, +AREA |
| Accessibility | +ACCESS | +INTDEN, +TRANS, +DRIVE, +RIDET |
| Distance to population center | -DISCBD | |
| MP timing | +TIME | |
| Regional influences | +SERCHG | +PERCHG, +POPDIF, +JOBCHG |

The maximum number of variables for any one equation in the recursive block is 12. The total number of exogenous variables in the simultaneous block is 29. Thus, the total number of instrumental variables for the two-stage least squares solution of the simultaneous block approaches the study sample size of 40. The result is that that the available degrees of freedom for model testing of any one simultaneous equation are 10 (=40-29-1). This exceeds the minimum of 6 degrees of freedom thought to be necessary to obtain meaningful results in the initial path analysis.

Note that the variables used in the initial model specification exclude four exogenous variables for which data could not be obtained (see Section III).These variables are DENCHG, ONLOT, SOILS, and ENERGY. Although these variables were not originally excluded from the initial model, it is instructive to include these changes at this point.

The theoretical basis for the initial model is discussed, by endogenous variable, in the following sections. Note that in these discussions the implicit assumption is made that all of the variables in a given equation are held constant while considering the effects of a particular independent variable. Also, although only the prime independent variable is discussed, the basis for corresponding alternative variables is identical.

### 1. RES Equation

RES = MANF + OFFICE + HIWAYS + REC
+ RECAP1 + LENGTH + TIME
- LIMITS + DUACRE + HSECHG
+ INCOME + JOBS + VALUE
- LAND - STAY + VACANT
+ RZONED + ACCESS

Residential development (RES) is hypothesized to respond to increased manufacturing (MANF) and office-professional services (OFFICE) activity due to employment opportunities, and to highway lane miles (HIWAYS), as such development (particularly high density residential) locates in areas well served by transportation networks. In addition, active, outdoor recreational development (REC) is included as a causal factor since vacation homes often locate near such facilities.

Three different wastewater major project variables (RECAP1, LENGTH, TIME) are included since wastewater facilities have strong growth-inducing impacts on residential development (both single family and high density).

Growth in regional housing stock (HSECHG) was included to control for the influence of regional growth on residential development in the area of analysis. The index of mobility (STAY) was included to control for regional shifts in population.

Five base year socio-economic indicators are included in the RES equation. Residential density (DUACRE) and employment density (JOBS) are both expected to influence total future residential development. New residential construction is attracted to areas with relatively high income (INCOME) and existing residential property values (VALUE), and low vacant land prices (LAND). Three separate variables for representing base year land use constraints are included in the RES equation. A constraint on residential development (particularly high density) is the existence of on-site wastewater disposal restrictions (LIMITS).

Since development is controlled to some extent by the zoning classifications set in an area, it is expected that the amount of land zoned for residential (RZONED) is causally related to residential development. In addition, a prime factor encouraging new residential development is the availability of vacant, developable land (VACANT). The density of limited-access highway interchanges (ACCESS) is a measure of an area's accessibility, and so can have an effect on the location of new residential construction, particularly on the fringe of large metropolitan areas.

## 2. COMM Equation

COMM = RES + OFFICE + HIWAYS + COMCHG
+ INCOME + OFFJOB + POPDEN
+ CZONED + ACCESS + TIME

Commercial development (COMM) is hypothesized to respond to residential (RES) development and business activities (OFFICE), as both of these stimulate demand for commercial services. The existence of a transportation network (HIWAYS) is also important in servicing demand, particularly with larger commercial developments, such as regional shopping malls.

Growth in regional retail-wholesale trade employment (COMCHG) was included to control for the influence of regional growth on commercial development in the area of analysis.

Three base year socio-economic measures are included in the COMM equation. Relative median income (INCOME) generally determines the amount of commercial expenditures and, hence, the amount of commercial development. Densities of population (POPDEN) and office employees (OFFJOB) are indicators of the potential market for commercial services.

The density of limited-access highways interchanges (ACCESS) is a measure of an area's accessibility, and so has a strong effect on the location of commercial development.

Since development is controlled to some extent by the zoning classifications set in an area, it is expected that the amount of land zoned for commercial use (CZONED) is a determinant of commercial development.

Finally, the amount of commercial development is thought to be related to the number of years available for secondary growth (TIME).

## 3. OFFICE Equation

$$\text{OFFICE} = \text{RES} + \text{COMM} + \text{MANF} + \text{HIWAYS} + \text{INCOME} + \text{OFFJOB} + \text{POPDEN} + \text{ACCESS} + \text{OZONED} + \text{TIME} + \text{OFFCHG} + \text{VACANT}$$

Office-Professional Services development (OFFICE) is hypothesized to respond to residential (RES) and manufacturing (MANF) development as these create the principal market for Office-Professional Services. The proximity to commercial development (COMM) and to transportation networks (HIWAYS) are strong locational factors in office building location.

Growth in regional office employment (OFFCHG) was included to control for the influence of regional growth on office-professional services development in the area of analysis.

Three base year socio-economic measures were included in the OFFICE equation. Office employment density (OFFJOB) is expected to increase total future office development. Densities of population (POPDEN) and office employees (OFFJOB) are indicators of the demand and personnel available for office-professional services.

The density of limited-access highway interchanges (ACCESS) is a measure of an area's accessibility, and so has a strong effect on the location of office-professional services development.

Since development is controlled to some extent by the zoning classification set in an area, it is expected that the amount of land zoned for office use (OZONED) is a determinant of office-professional services development. In addition, a prime factor encouraging new office development is the availability of vacant, developable land (VACANT).

Finally, the amount of office development is thought to be related to the number of years available for secondary growth (TIME).

### 4. MANF Equation

$$\begin{aligned} \text{MANF} = {} & \text{HIWAYS} + \text{RECAP1} + \text{LENGTH} + \text{MANJOB} \\ & + \text{RRMILE} + \text{ACCESS} + \text{VACANT} + \text{IZONED} \\ & - \text{LIMITS} + \text{TIME} - \text{LAND} + \text{MANCHG} \end{aligned}$$

Prime location requirements for manufacturing industries (MANF) include proximity to good highways (HIWAYS, ACCESS) and availability of railroad access (RRMILE).

Three different wastewater major project variables are included since the placement (LENGTH), sizing (RECAP1), and timing (TIME) of wastewater facilities effects the location of industrial facilities (MANF).

Growth in regional manufacturing employment (MANCHG) was included to control for the influence of regional growth on manufacturing development in the area of analysis.

Two base year socio-economic measures are included in the MANF equation. Manufacturing employment density (MANJOB) is expected to increase total future manufacturing development, while vacant land costs are expected to decrease such development.

Three separate variables for representing land use constraints are included in the MANF equation. A constraint on industrial development is the existence of on-site wastewater disposal restrictions (LIMITS). Since development is controlled to some extent by the zoning classifications set in an area, it is expected that the amount of land zoned for industrial use (IZONED) is causally related to manufacturing development. In addition, a factor encouraging new manufacturing development is the availability of vacant, developable land (VACANT).

### 5. HIWAYS Equation

$$
\begin{aligned}
\mathrm{HIWAYS} = {} & \mathrm{RES} + \mathrm{COMM} + \mathrm{OFFICE} + \mathrm{MANF} \\
& + \mathrm{DUACRE} + \mathrm{JOBS} + \mathrm{DRIVE} \\
& - \mathrm{DISCBD} + \mathrm{TIME} - \mathrm{INTDEN}
\end{aligned}
$$

Non-expressway highway lane mile development (HIWAYS) is hypothesized to respond to residential (RES), commercial (COMM), office-professional services (OFFICE), and manufacturing (MANF) developments since each of these generate substantial vehicular traffic and highway facilities service this transportation need.

The relative limited access highway interchange density of the area of analysis (INTDEN) is expected to restrict the development of non-expressway highway development since these two types of highway service many of the same needs. Highway facilities provide access to population and employment centers, such as central business districts. Thus, an inverse relationship between the distance to the CBD (DISCBD) and highway development is expected.

Two base year socio-economic indicators are included in the HIWAYS equation. Residential density (DUACRE) and employment density (JOB) are both expected to increase future highway construction.

The regional density of motor vehicle drivers (DRIVE) is considered to be a further indicator of the demand for highway facilities.

Finally, the amount of highway development is thought to be related to the number of years for secondary growth (TIME).

### 6. EDUC Equation

$$
\begin{aligned}
\mathrm{EDUC} = {} & \mathrm{RES} + \mathrm{DUACRE} + \mathrm{VALUE} + \mathrm{KIDS} + \mathrm{RZONED} + \mathrm{TIME} \\
& - \mathrm{STAY} + \mathrm{HSECHG}
\end{aligned}
$$

Educational development (EDUC) is hypothesized to respond to residential development (RES) since this housing stock generally determines the market for educational facilities.

Growth in regional housing stock (HSECHG) was included to control for the influence of regional growth on educational development in the area of analysis. An index of mobility (STAY) was included to control for regional shifts in population.

Three base year socio-economic measures are included in the EDUC equation. Public school enrollment density (KIDS) is expected to increase the total future educational facility development. Residential density (DUACRE) and the value of existing residential property (VALUE) are both expected to increase the amount of total future educational facilities.

Since development is controlled to some extent by zoning classifications set in an area, it is expected that the amount of land zoned for residential (RZONED) is causally related to future educational development.

Finally, EDUC is thought to be related to the number of years available for secondary growth (TIME).

### 7. REC Equation

$$\begin{aligned} \mathrm{REC} = {} & \mathrm{RES} + \mathrm{EDUC} + \mathrm{INCOME} + \mathrm{KIDS} \\ & - \mathrm{LAND} + \mathrm{RZONED} + \mathrm{POPDIF} \end{aligned}$$

Active, outdoor, recreational development (REC) is hypothesized to respond to residential (RES) development due to the demand it creates for recreational opportunities. Educational facilities (EDUC), due to their youth orientation, also generate significant demand for active, outdoor recreational facilities.

Growth in total regional population (POPDIF) was included to control for the influence of regional growth on recreational facilities in the area of analysis.

Three base year socio-economic measures are included in the REC equation. Relative median income (INCOME) and public school enrollment density (KIDS) are determinants of the amount of active, outdoor recreation in a community. To a lesser extent, the cost of vacant land (LAND) determines the size and numbers of recreational areas. The amount of land zoned for residential (RZONED) also gives an indicator of the future demand for recreational areas.

### 8. WHOLE Equation

$$\begin{aligned} \text{WHOLE} = {} & \text{COMM} + \text{MANF} + \text{HIWAYS} + \text{ACCESS} \\ & + \text{WWJOB} - \text{LAND} + \text{POPDEN} \\ & + \text{VACANT} + \text{IZONED} + \text{TIME} + \text{COMCHG} \end{aligned}$$

Wholesale-warehousing development (WHOLE) is hypothesized to respond to related manufacturing (MANF) and commercial (COMM) developments. A locational factor which attracts wholesale-warehousing is accessibility to highways (HIWAYS, ACCESS).

Growth in regional retail-wholesale trade employment (COMCHG) was included to control for the influence of regional growth on wholesale-warehousing development in the area of analysis.

Three base year socio-economic indicators are included in the WHOLE equation. Densities of population (POPDEN) and wholesale/warehousing employees (WWJOB) are indicators of the market and personnel for new wholesale/warehousing development. Due to the large land area of such developments, they are very sensitive to the cost of vacant land (LAND).

Prime factors encouraging new wholesale/warehousing are the amount of land zoned for industrial use (IZONED) and the amount of vacant developable land (VACANT).

Finally, WHOLE is thought to be related to the number of years available for secondary growth (TIME).

### 9. OTHER Equation

$$\begin{aligned} \text{OTHER} = {} & \text{RES} + \text{OFFICE} + \text{MANF} + \text{REC} \\ & + \text{SERCHG} + \text{INCOME} + \text{JOBS} + \text{UNIV} \\ & - \text{LAND} + \text{ACCESS} - \text{DISCBD} + \text{TIME} \end{aligned}$$

Other urban development types (OTHER) include religious, cultural-entertainment, and hotel-motel activities. These are hypothesized to respond to residential development (RES) as this is the principal source of demand for the services of other urban development types. In addition, office-professional services (OFFICE), manufacturing (MANF) and active, outdoor recreational (REC) developments are causal factors since each can generate overnight visitors and, hence, demand for hotel-motel services.

Growth in regional services employment (SERCHG) is included to control for the influence of regional growth on other urban development types in the area of analysis.

Three base year socio-economic measures are included in the OTHER equation. Since other urban development includes hotel-motel land uses and, since employment opportunities provide a market for hotel-motel services, total employment density (JOBS) is expected to increase total future other urban development. The existence of a university or college (UNIV) in the area is expected to increase the development of cultural activities to service the academic community and to increase hotel-motel development through accommodation requirements for professionals and student's families. All of these are elements of the variable OTHER. The types of development including OTHER are also closely related to relative income levels (INCOME).

Since cultural, religious and hotel-motel activities all service, and hence locate, near population and employment centers, an inverse relationship between the distance to the nearest central business district (DISCBD) and other urban development is expected. The density of limited-access highway interchanges (ACCESS) is a measure of an area's accessibility, and so can have an effect on the location of new cultural-entertainment and hotel-motel activities.

The cost of vacant land (LAND) is often a prime factor in the location of OTHER developments.

Finally, OTHER is thought to be related to the number of years available for secondary growth (TIME).

## III. PHASE II - DATA COLLECTION

The principal objective of the second phase was to collect a sufficiently large and diverse cross-sectional data base on which to develop the causal and predictive model.

### A. CASE STUDY SELECTION

Due to the critical importance of data quality and the large manpower effort required to collect the data base for this study, the first task was to carefully select a set of 40 projects for case studies. These case studies were wastewater major projects, distributed nationwide, which had the potential, upon construction, for inducing a significant quantity of land development in their communities. Identification of the case study projects was based primarily upon an extensive mail survey of municipalities and regional planning agencies, supplemented by data obtained from a review of secondary (published) data sources.

The approach to the selection of case study major projects involved three steps. First, a list of selection criteria were developed. Second, these criteria were applied to information derived from primary and secondary sources to obtain a preliminary list of 46 potential case studies. Finally, detailed inquiries were made regarding each potential case study to verify its compliance with every one of the selection criteria, and a final set of forty case study major projects compiled. These efforts are reported on in the sections below.

#### 1. Selection Criteria

The selection criteria were designed to standardize certain aspects of the model data in order to avoid difficulties in the analysis and to minimize the overall effort of the data collection program. They are summarized as follows:

##### a. Geographical Distribution Requirements

These were specified to ensure a generalized model applicable nationwide. In summary:

- All cases had to be located in the continental United States.
- At least two cases had to come from each federal region.
- No more than five cases could come from any one state.
- No more than two cases could come from any one SMSA*, and
- The area of analysis of any two cases could not be contiguous or overlapping.

b. Size, Type and Timing Requirements

These were the specified by the definition of a wastewater major project, as repeated below:

- A wastewater major project is considered to involve principally the construction or extension of interceptor or collector sewer lines during the period 1958 to 1962+ in a community in the United States. If construction date information is not available, then a grant funding date in the period 1956-1960 is acceptable. The project had to affect an increase in absolute system collection capacity of 1 MGD or more, and had to cost a minimum of $200,000 to construct. Phased projects are considered if the first phase of construction on the collection network was complete within 1958-1962+ and the last phase of construction was complete by 1965.

c. Potential for Induced Growth Requirements

These were specified to ensure that induced development could have occurred in connection with a wastewater project, and are given by the definition of area of analysis, as repeated below:

- The area of analysis is defined as the legal service area of a wastewater major project in the base year. It must be a minimum size of 5,000 acres (or approximately eight square miles), and contain significant amounts of vacant developable land, some of which must be more than 5,000 feet from the nearest interceptor line in the base year.

---

*Standard Metropolitan Statistical Area.

+It was necessary to extend this time period to 1963 in order to obtain 40 case study wastewater projects.

### d. Availability of Data Requirements

These ensured that the municipality, regional planning agency or water district associated with each case study project had sufficient records for the project to estimate values for important exogenous model variables. Specifically, the following data were required:

- A map of the legal service area showing the extent of the wastewater collection network in the base year
- Base year population in the legal service area
- Actual peak wastewater flow in the wastewater system in the base year
- Total hydraulic design capacity of the major project (both collection network and treatment plant).

## 2. Information Sources

Case study major projects were identified from both primary and secondary data sources, as discussed below.

### a. Primary Data Sources

The primary source of information for case study selection was a nationwide mail survey. Since the base year of interest is 1960, it is expected that many of the sewer project records, which are almost two decades old, may be lost or destroyed. For this reason, the survey effort was directed at two possible sources of information: (1) those municipalities in which the project is located, and (2) regional planning agencies.

The first survey consisted of regional planning agencies, the best source of case study projects for GEMLUP-I. The 1977 Directory of such agencies, obtained from the National Association of Regional Councils [118], was used to identify the agencies. Due to the large number of such groups (669) we could not survey them all. Therefore, within each federal region, one or two representative states were chosen for the survey, and only regional planning agencies which were also A95 Clearinghouses (the local agency responsible for overseeing federally funded projects) were contacted. The total number of agencies surveyed is 213; relevant statistics associated with the survey are summarized by federal region in Table 3-1.

The second survey was of municipal department of public works in the selected States. EPA's Project Register [52] of over 15,000 wastewater projects nationwide receiving matching federal funds between 1956 and 1973 was screened using the definition of major project to obtain the survey list. A sample of the mail survey questionnaire is shown in Table 3-2. The total number of municipalities surveyed was 291. Relevant statistics associated with this survey are summarized by federal region in Table 3-1.

The results of the case study mail survey (see Table 3-1) show a total of 39 case study wastewater projects identified by this approach. Note that it was necessary to extend the time period for construction of a major project to 1963 to obtain a sufficient number of case studies. Of the 39 selected, 9 or 23% finished construction in 1963. A list of the selected case study major projects is given in Table 3-3.

### b. Secondary Data Sources

As part of the literature search performed in Phase I of this study, an exhaustive search for published case studies was undertaken. A review of secondary data with regard to the case study selection criteria produced only seven additional case study major projects, these are numbers 27,36-40, and 46 in Table 3-3 [101, 110-112]. Most of the potential projects identified in secondary sources were inappropriate due to noncompliance with one or more of the selection criteria. The remainder did not present sufficient data to judge the appropriateness of a project. This does not mean that the secondary data sources were unessential to the selection process. In fact, they were invaluable in providing data on the wastewater projects identified through the mail survey which allowed the selection of the case studies. Sources that proved to be particularly useful were the following:

## TABLE 3-1

## SUMMARY OF SURVEY QUESTIONNAIRE RETURNS

| Federal Region | Number Mailed | Number Returned | Number Identifying Projects | Number of Case Study Projects |
|---|---|---|---|---|
| Regional Planning Councils | | | | |
| I | 26 | 4 | 1 | 2 |
| II | 3 | 1 | 1 | 1 |
| III | 26 | 4 | 2 | 1 |
| IV | 27 | 2 | 1 | 0 |
| V | 30 | 7 | 2 | 2 |
| VI | 24 | 4 | 0 | 0 |
| VII | 20 | 1 | 0 | 0 |
| VIII | 9 | 0 | 0 | 0 |
| IX | 18 | 1 | 0 | 0 |
| X | 30 | 4 | 1 | 0 |
| Subtotal | 213 | 28 | 8 | 6 |
| Municipalities | | | | |
| I | 39 | 10 | 9 | 4 |
| II | 21 | 3 | 3 | 1 |
| III | 31 | 6 | 4 | 2 |
| IV | 52 | 12 | 8 | 7 |
| V | 39 | 10 | 9 | 2 |
| VI | 34 | 8 | 6 | 5 |
| VII | 20 | 2 | 2 | 2 |
| VIII | 9 | 1 | 1 | 1 |
| IX | 35 | 19 | 14 | 6 |
| X | 11 | 5 | 4 | 3 |
| Subtotal | 291 | 76 | 60 | 33 |
| Total | 504 | 104 | 68 | 39 |

Return rate = 21%
% of returns identifying projects = 65%
% of total identifying projects = 14%

## TABLE 3-2

## SAMPLE SURVEY QUESTIONNAIRE

**Name of Respondent:** **Telephone #:**

**Organization:**

**Address:**

**Name of Project:** **Project #:**

**Start/Completion Dates:**

| Location of Project | | | Circle One |
|---|---|---|---|
| Municipality | - Name: | Size: | Acres<br>or<br>Sq. Miles |
| Drainage Basin or Sewershed | - Name: | Size: | |
| Approximate Size of Interceptor Service Area: | | | Acres or Sq. Miles |

Design Excess Capacity of Interceptor Main: %

Available Data Sources in Area for Planning Information

3C Planning Agency:

Detailed Land Use Inventory or Aerial Photographs for Municipality circa 1970:

Please attach a map of any scale of the area encompassing the interceptor project. Feel free to make any pertinent comments concerning the project below. If you have any questions, call Peter Guldberg at 617-657-4250 X119. Thanks for your assistance.

## TABLE 3-3
## CASE STUDY MAJOR PROJECTS

| ID Number | Project Name | Location | ID Number | Project Name | Location |
|---|---|---|---|---|---|
| 1 | Willimantic SS* | Willimantic, CT | 19 | Lick Creek Interceptor | Danville, IL |
| 2 | Leeds and Ryan Road Interceptors | Northampton, MA | 20 | Roseville Interceptor | Roseville, MN |
|  |  |  | 21 | Denton SS | Denton, TX |
| 3 | Woodhaven and Vinebrook Sewers | Lexington, MA | 22 | Lamesa SS | Lamesa, TX |
| 4 | Hudson SS | Hudson, MA | 23 | Garland SS | Garland, TX |
| 5 | Milford SS | Milford, MA | 24 | Coldwater Creek Interceptor | St. Louis County, MO |
| 6 | Eatontown SS | Eatontown, NJ | 25 | Liberty West Side Sewers | Liberty, MO |
| 7 | Southside Interceptor | Plainfield, NJ | 26 | Florence SS | Florence, CO |
| 8 | Cox Creek SS | Anne Arundel County, MD | 27 | Boulder SS | Boulder, CO |
| 9 | Division O-22nd Street Interceptor | Richmond, VA | 28 | Casa Grande SS | Casa Grande, AZ |
| 10 | Western Branch SS | Portsmouth, VA | 29 | South Phoenix Interceptor | S. Phoenix, AZ |
| 11 | Crane Creek SS | Melbourne, FL | 30 | South Interceptor | Vallejo, CA |
| 12 | Lakeland SS | Lakeland, FL | 31 | Demair Interceptor | Turlock, CA |
| 13 | Southeast Interceptor | Clearwater, FL | 32 | Vacaville SS | Vacaville, CA |
| 14 | Tar River Interceptor | Greenville, NC | 33 | Vancouver SS | Vancouver, WA |
| 15 | Upper Walnut Creek Interceptor | Raleigh, NC | 34 | North and South Interceptors | Auburn, WA |
|  |  |  | 35 | Everett SS | Everett, WA |
| 16 | Asheboro SS | Asheboro, NC | 36 | Bennington SS | Bennington, VT |
| 17 | McMullen Creek SS | Charlotte, NC | 37 | Jamestown SS | Jamestown, NY |
| 18 | Effingham SS | Effingham, IL | 38 | Farmington SS | Farmington, NM |

* Sewer System

TABLE 3-3 (Continued)

CASE STUDY MAJOR PROJECTS

| ID Number | Project Name | Location | ID Number | Project Name | Location |
|---|---|---|---|---|---|
| 39 | Vista/Carlsbad SS | Vista, CA | 43* | Orange SS | Orange, TX |
| 40 | Moscow SS | Moscow, ID | 44* | Texarkana SS | Texarkana, TX |
| 41* | South Amherst Interceptor | Amherst, MA | 45* | Northern Interceptor | Livermore, CA |
| 42* | Richfield/Edina Interceptor | Richfield/Edina, MN | 46* | Greensboro SS | Greensboro, AL |

* Not selected as one of the final 40 case study major projects

- The Accelerated Public Works Directory for 1964 [111] and the EPA Wastewater Contract Awards computer data file, maintained by the EPA Municipal Construction Division [101], both of which provided wastewater completion dates for all types of wastewater projects funded under PL-87-658 and PL-84-660.
- The EPA 1968 Inventory, Municipal Waste Facilities [112] which tabulates all municipal waste facilities installed by 1968, including non-federally funded projects. This publication provided data on completion date, drainage basin, and plant design capacity and initial average wastewater flow.

In addition to published data sources, the EPA Municipal Construction Division supplied information [113] on two wastewater facilities that might be potential case study projects. The first of these involved three sewage treatment plants for Lancaster, PA. Of these three projects, only one was constructed in the time period of interest and this one unfortunately involved only the upgrading of the treatment plant and not an extension of the collection network. The second set of facilities resides in Chattanooga, TN and involves two sewage treatment plants. Unfortunately, neither of these projects was associated with an expanded collection network until 1964 when construction on interceptor extension began. Thus, no case study projects were selected from these data.

### 3. Final Selection

Table 3-3 lists data for wastewater exogenous model variables which had to be known to determine whether a given project satisfied the selection criteria for a case study. Since the preliminary list of 46 projects drawn from primary and secondary sources did not provide all such data, a telephone survey of municipal departments of public works, regional planning agencies and state soil conservation service offices was carried out to complete the data shown in Table 3-3. In many instances values for the variables PEAK and CAPAC1 could not be obtained and so estimates were used instead. In these instances, the peak flow in a major project's collection system was estimated by scaling a known average flow by the population-dependent peaking factors given in Figure 2-6 of reference [100]. When the hydraulic design capacity of the collection network had to be estimated, one of two different techniques was employed. The most common estimation procedure was to calculate CAPAC1 as follows:

$$CAPAC1 = PEAK * (1 + RECAP1/100)$$

The second method used was to sum estimated hydraulic flow rate for each major pipe in the collection system. Hydraulic flow was determined from nomographs of Manning's formula [100] which required pipe diameter and slope. In the absence of data for the latter parameter, average values for hydraulic flow based on pipe diameter were used [114].

The telephone survey revealed that three major projects should be dropped from consideration as case studies, since they no longer satisfied the definition of major project. The first of these, #41 in Amherst, MA, only has a design excess capacity in the base year of 5%, i.e., there was essentially no excess capacity in the collection system. The second, #42 in Richfield, MN, was completed in 1954, not 1961 as previously thought. The last project was #46 in Greensboro, AL. Since this area has never been surveyed by the U.S. Geological Survey, no maps were available and it would have been very difficult to define an area of analysis and collect data for this case. At the Project Officer's direction, three additional case study projects were eliminated (#45 in Livermore, CA; #43 in Orange, TX; and #44 in Texarkana, TX) to form the final set of 40 case study major projects.

## B. DATA COLLECTION

### 1. Transportation Data

Prior to the data collection effort, the GEMLUP-I VMT model [19] was reviewed to identify what data were required to test and validate the model. Thus, the purpose of this task was to determine the most appropriate spatial extent of data sources; the best available agency sources of data; and to initially select variables for which data were to be collected for GEMLUP-II. In performing this task, an initial choice of VMT case studies was made.

### a. Spatial Extent of Data Sources

The basic question to be dealt with concerns the most appropriate transportation planning geographic unit for analyzing and aggregating data. The practice in transportation planning is to use either traffic districts or traffic zones. Traffic districts are relatively large spatial units normally utilized in projecting regional travel. Traffic zones, on the other hand, are spatial units roughly corresponding with census tracts. In examining each of these traffic data units, it was kept in mind that the range in size of GEMLUP-II study areas would average from 20,000 to 40,000 acres with some study areas as small as 5,000 acres. A second factor which was considered was that it would necessarily be easier to aggregate data from smaller traffic data units that it would be to disaggregate data from larger units; this consideration is important in maintaining flexibility in the general VMT model.

In considering the use of data from traffic districts versus those from traffic zones, it was concluded that traffic zones would best serve the purposes of GEMLUP-II. First, traffic zones are more to scale with the selected study areas. Second, traffic zone population and economic data are more detailed and correspond quite closely with Census Tracts. Third, the number of traffic zones per study area is manageable with regard to the volume and type of data they would generate. (For example, a recent study of a 38.8 square mile Southern California city of 24,832 acres contained 86 traffic zones [115]). Lastly, it is generally agreed that some reliable transportation models are built upon zonal data [116].

All of these factors clearly point to the conclusion that for GEMLUP-II, the utilization of traffic zone data are far superior to the generalized--and often incorrect--data that could be obtained from traffic districts.

### b. Agency Sources of Data

The next basic questions associated with approaches to the VMT model validation was the best agency source of data. It was initially suggested that agencies associated with the so-called 3-C process be utilized. These are transportation agencies operating under the Federal-Aid Highway Act of 1962 which seeks to develop transportation programs based on a "continuing, comprehensive transportation planning process carried on cooperatively in the state and local communities [44]." There are, however, several glaring problems associated with using data from 3-C agencies. First, the process is applicable only to areas with a population of 50,000 inhabitants or more; the majority of VMT case studies initially selected have a population below this figure. Second, unnecessary difficulty would be encountered in disaggregating informmation from 3-C agencies as to be applicable to sub-areas covered by the case studies; this is verified by virtue of the fact that most 3-C agencies are State Highway Departments and their data commonly relate to fairly large regional areas. Finally, experience has shown that 3-C agency data vary widely in its accuracy and detail; no two 3-C agencies studies follow identical methods or procedures; and the process outlined in the Federal Highway Administration (FHWA) Policy and Procedure Memoranda, PPM 50-9 are not always rigorously followed [117].

A more appropriate source of data was Regional Planning Agencies covering areas in which the case studies are located. Of the case studies initially selected for transportation data collection, all were within an area covered by a Regional Planning Agency. These planning agencies were founded at a point in time where it is highly likely that they would have developed data under the appropriate study period of 1970 [118]. These Regional Planning Agencies were also likely to have a finer level of data dealing with transportation, population, and socio-economic characteristics.

It should be noted, however, that a given Regional Planning Agency could not have all the data needed for validation of the VMT model. In such a case, additional sources had to be contacted for data. These included a

City or County Traffic Engineer for data dealing with hierarchy of streets, traffic volume and design capacity, and vehicle speed. Also, County and City Planning Departments had information on population, socio-economics, and land use which were particularly useful in addressing trip productions and attractions on a zonal basis.

### c. Initial Choice of VMT Case Studies

Of the GEMLUP-II case studies listed in Table 3-3, 11 were selected for transportation data collection. Data from these 11 areas were used in the validation of the VMT model. They are:

VMT Case Studies

Lexington, Massachusetts (test case)
Richmond, Virginia
Charlotte, North Carolina
Clearwater, Florida
Roseville, Minnesota
St. Louis, Missouri
Boulder, Colorado
S. Phoenix, Arizona
Vallejo, California
Auburn, Washington
Bennington, Vermont

The criteria associated with the selection of these VMT case studies were geographic representativeness, urban development pattern, socio-economic characteristics, and the existence of a Regional Planning Agency under the study period. These case studies range in population (popcom) from 1,000 to 201,564. In this way, a full range of cases were available for testing the VMT model and determining the overall validity of it as a generalized model.

### d. Initial Choice of Variables

On the basis of the literature search and analysis of that literature, there were numerous variables which respective authors found useful in their specific transportation studies. These variables differed according to the particular purpose of their studies [116,119-123]. For example, Table 3-4 lists and ranks those variables which the U.S. Department of Transportation

## TABLE 3-4

## SELECTED VARIABLES FOUND SIGNIFICANT IN URBAN TRANSPORTATION PLANNING

| | | |
|---|---|---|
| A. | VARIABLES FOUND SIGNIFICANT IN ZONAL TRIP GENERATION | |
| 1. | Demographic Data | |
| a. | Total Population | 1* |
| b. | Age, Sex, Race, etc. | 3 |
| c. | Number of Household Units | 1 |
| d. | School Enrollment | 2 |
| e. | Family Life Cycle | 3 |
| 2. | Economic Data | |
| a. | Total Employment (1)** | 1 |
| b. | Selected Employment (2) | 1 |
| c. | Employment by Industry (3) | 3 |
| d. | Employment by Residence (4) | 1 |
| e. | Labor Force (5) | 3 |
| f. | Labor Force by Occupation and Industry (6) | 3 |
| g. | Median Income | 1 |
| h. | Income Stratified | 3 |
| i. | Automobile Ownership | 1 |
| j. | Dwelling Units without Automobiles | 2 |
| k. | Retail Sales | 2 |
| l. | Average Home Value | 3 |
| 3. | Land Use Data | |
| a. | Specific Activities | 3 |
| b. | Selected Categories | 1 |
| B. | VARIABLES FOUND SIGNIFICANT IN DWELLING UNIT TRIP GENERATION | |
| a. | Car Ownership | 1 |
| b. | Family Size | 1 |
| c. | Number of Persons 5 Years old and over in Household | 1 |
| d. | Length of Residency | 3 |
| e. | Family Income | 2 |
| f. | Number of Persons 16 years old and over | 1 |
| g. | Number of Persons 16 years old and over who drive | 1 |
| h. | Age of the Head of Household | 2 |
| i. | Distance from CBD | 3 |
| j. | Stage in Family Life Cycle (7) | 1 |
| k. | Occupation of Head of Household | 1 |
| l. | Structure Type | 1 |

---

* Key to Weights: 1 = Essential Data, 2 = Desirable Data, 3 = Useful Data.

** Number referenced to notes appearing on the next page.

Source: [116]

## TABLE 3-4 (CONTINUED)

## SELECTED VARIABLES FOUND SIGNIFICANT IN URBAN TRANSPORTATION PLANNING

(1) *Total Employment*. All persons employed in the study area, or employment by place of employment.

(2) *Selected Employment*. An employment grouping for the study area that classifies workers by occupation, industry, or class-of-workers.

(3) *Employment by Industry*. All employment in the study grouped by SIC, or employment by place of employment or industry.

(4) *Employees by Residence.* All persons living in the study area who are employed someplace.

(5) *Labor Force.* All persons living in the study area who are either employed or unemployed.

(6) *Labor Force by Industry*. All persons living in the study area who are either employed or unemployed, and grouped according to industry of employment or experience of unemployment.

(7) *Stage in the Family Life Cycle*. Based on age of the head of the household, marital status, and age and number of children.

found significant in transportation planning. In a study performed by the Michigan Department of State Highways, the following variables were determined to be significant in predicting trips at the zonal level [123]:

1. Total Population
2. Number of Dwelling Units
3. Autos Available
4. Resident Labor Force
5. Total Employment
6. Manufacturing Employment
7. Government Employment
8. Retail and Wholesale Employment
9. Service Employment
10. Selected Employment
11. Retail Sales

Finally, a nationwide survey [116] ranked the importance of seven household characteristics for trip generation as follows:

| Variable | Rank | Regression Beta Coefficient |
|---|---|---|
| Family Size | 1 | .29 |
| Car Ownership | 2 | .23 |
| Income | 3 | .14 |
| Stage in Family Life Cycle | 4 | .13 |
| Occupation of Head of Household | 5 | .11 |
| Density of Neighborhood | 6 | .10 |
| Distance from CBD | 7 | Insignificant |

As is self-evident from the above, the choice of variables differs from study to study. These lists and tables reflect the "biases" of individual studies. In selecting variables for validating the overall GEMLUP VMT model, as well as trip generation and trip length characteristics, the following three interrelated factors must be considered:

1. Pattern of land use (location and intensity);
2. Social and economic characteristics of the study area; and
3. The type and extent of transportation facilities in the study area.

In examining these factors within the context of the overall GEMLUP VMT model, the following variables were selected as being most appropriate for the validation exercises:

1. Pattern of Land Use
    a. Selected Categories of Land Use (acres, square feet) by zone (tabulation) and by map

2. Social and Economic Characteristics
    a. Total Population by Zone
    b. Total Employment by Zone
    c. Selected Employment by Occupation, Industry, or Class-of-Workers by Zone
    d. Number of Households (Dwelling Units Density) by Zone
    e. Median Income by Zone
    f. Stratified Income by Zone
    g. Automobile Ownership by Zone (Registration and Fleet Mix--car, truck, bus)

3. Transportation Facilities
    a. Map of Transportation Network (local, collector, arterial, expressway)
    b. Traffic Volume (ADTs) and Design Capacity for Each Major Roadway
    c. Average Trip Lengths for Each Major Roadway
    d. Number of Interzonal Trips by Purpose per Day (home, work, shopping, social-recreation, school, personal business)
    e. Vehicle Speed for Each Major Roadway
    f. Zonal Trip Generation Rates by Land Use Activity
    g. Zonal VMT

These variables are specifically identified as an assurance that all necessary information for VMT model validation will be obtained. The rationale behind this approach is that the data, methods, and procedures employed in areas throughout the country vary considerably; for example, one cannot be confident that information on trip generation rates are necessarily available nor that this information was calculated with accuracy. Thus, it may be necessary to estimate trip rates for a given purpose by dividing the number of interzonal trips (3d) by zonal land use (1a). Similarly, zonal VMT can be estimated by summing over all major roadways the product of ADT (3b) and roadway length (3a). The initial choice of VMT variables, therefore, may represent data that will not be needed for each VMT model case study, but it will ensure that enough data are available to guarantee overall consistency in validation. In addition, this data will be invaluable in modifying the existing GEMLUP VMT model to incorporate more detailed considerations if the validation indicates a need for such revision.

### 2. Survey Forms

The list of data requirements developed in Phase I, the list of data requirements for the GEMLUP-I VMT model validation, and the list of 40 case study wastewater major projects defined the data collection program. These requirements were formalized in a set of survey forms. Survey forms for the collection of data from all sources but U.S. Census publications are reproduced in this section as Exhibit 3-1. Census data collection was performed by the Project Officer and coded directly onto keypunch forms. Questions on soil suitability for buildings and septic tanks were included on the regional planning agency form after discussions with the U.S. Soil Conservation Service revealed that their published soil data are available for only a small fraction of the land area in the U.S. Planning agencies were expected to have access to a larger set of unpublished soil data.

OMB Clearance
No. 158-S76007

EXHIBIT 3-1

# REGIONAL PLANNING AGENCY SURVEY FORM

## APPOINTMENT SCHEDULE

Dates ____________________________________________

Time ____________________________________________

Name ____________________________________________

Organization ____________________________________

Address __________________________________________
__________________________________________
__________________________________________

Telephone Number ________________________________

EXHIBIT 3-1

## REGIONAL PLANNING AGENCY SURVEY FORM

Name of Project ______________________________

Year Completed (t) __________

Community Served ______________________________

County ____________________

River Basin ____________________

Area of Analysis ____________________ acres

USGS MAP AND ROAD ATLAS DATA (Try to compute these before visit to RPA)

| | | |
|---|---|---|
| 1. | Distance in miles from centroid of area of analysis to centroid of nearest CBD (majority over 100,000 population) in year t | ☐ |
| 2. | Railroad mileage in area of analysis in year t | ☐ |
| 3. | Distance in miles from centroid of area of analysis to centroid of nearest commercial airport | ☐ |
| *4. | Number of limited-access highway interchanges in area of analysis in year (t+5) | ☐ |
| *5. | Number of limited-access highway interchanges in the county in year (t+5) | ☐ |
| *6. | Non-limited-access highway lane miles in area of analysis in 1970 (use worksheet) | ☐ |

SPECIFIC QUESTIONS (Show them map of area of analysis)

| | | |
|---|---|---|
| 1. | Did a college or university exist in the area of analysis in 1960? | ☐ |
| 2. | What was the median price ($) of residential vacant land in the area of analysis in 1960?<br>(Alt. source: Call local realtors, real estate appraisers, visit county recorder for real estate sale prices) | ☐ |
| 3a. | How many passenger commuter rail stops were there in the area of analysis in year t?<br>(Alt. source: Call local transit authority) | ☐ |
| b. | How many local bus stops (transit system) were there in the area of analysis in year t?<br>(Alt. source: Call local bus company) | ☐ |
| c. | Sum (3a) and (3b) | ☐ |

---

*Verify with RPA the year of construction of interstate highways shown on USGS maps or road atlas and count only those interchanges in service in year (t+5) and those lane miles in service in 1970.

## EXHIBIT 3-1

4a. How much vacant developable acreage existed in area of analysis in year t?

b. How much vacant undevelopable acreage existed in area of analysis in year t?
(Alt. source: Land use map for 1960)

5. How many acres in the area of analysis were zoned for the following uses in year t?

Residential

Commercial

Office

Industrial

6a. Which of the following is true with regard to restrictions on residential on-lot sewage disposal for the time period t to 1970? (Note: If more than one applies, code largest number)

4 = on-lot disposal prohibited entirely

3 = on-lot disposal prohibited except on large lots

2 = on-lot disposal permitted but percolation tests are required

1 = on-lot disposal permitted but package plants prohibited

0 = no restrictions

b. What year did on-lot disposal restrictions go into effect?
(Note: If no restrictions, code 1970; if year is less than 1958, code 1958)
(Alt. source: Municipal health dept., DPW or sewer authority

7a. During the period t to 1970, were any government policies in effect to limit the number of hookups to the sewage system in the area of analysis?

1 = Yes

0 = No

b. What year did such policies take effect?
(Note: If no policies, code 1970; if year is less than 1958, code 1958)
(Alt. source: Municipal DPW or sewer authority)

## EXHIBIT 3-1

8. How much active, outdoor recreational acreage was there in the area of analysis in 1970? This is defined to include:

   - Golf courses, driving ranges, miniature golf
   - Tennis courts
   - Skating rinks
   - Riding stables
   - Ski and tobaggan slopes
   - Athletic fields
   - Play lots and playgrounds
   - Swimming beaches and community pools
   - Yacht clubs, marinas and boating access sites
   - Dude ranches

   This is defined to exclude passive open space portions of large reservations, arboretums, parks and conservation areas such as state/national forests, wildlife refuges.

9. The following questions are required only if part of the area of analysis is "unblocked" (i.e., not covered by census tracts or blocks). In such cases, answer these questions only for the unblocked portions of the area of analysis.

   a. Number of dwelling units in 1970 - take off of aerial photos

   b. What is the proportion of multifamily dwelling units (3 or more households) in low rise and high rise structures?
      % Low
      % High

   c. What is the proportion of single family dwelling units in detached or attached (duplex) structures?
      % Detached
      % Attached

10. Soil suitability data for area of analysis-acreage with a "severe" soil type classification with regard to"

    a. Urban development (building sites)

    b. On-lot sewage disposal (septic tanks)

    (Alt. source: State Office of U.S. Soil Conservation Service - see attached list).

EXHIBIT 3-1

## HIWAYS WORK SHEET

| Link Name | Length (miles) | # of Lanes | Lane Miles |
|---|---|---|---|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

EXHIBIT 3-1

## IN-HOUSE DATA COLLECTION FORM

Community Served by Major Project________________________________

Year of Completion (t)_________

Year of Federal Funding_________

USGS MAPS

1. Using available watershed boundary and wastewater collection network maps for the major project, "draw" the boundaries of the area of analysis on USGS 1: 24,000 quadrangle maps of the area. The following definitions apply:

   The area of analysis is defined by the watershed boundaries of the area that drains, by gravity, to any point of the wastewater collection network, plus any areas outside the watershed connected by pumping stations and force mains. In the case of essentially flat terrain (e.g., river deltas), the area of analysis is restricted to the locus of points within 1,000 feet of any interceptor or trunk sewer line.

2. Compute the area of analysis in acres (refer to list of case study projects as the variable AREA has already been computed for some projects). ☐

3. Enter the name of the river basin which makes up most of the area of analysis. ________________________

4. Compute the length of interceptor sewer lines (in miles) transversing relatively undeveloped land (< 1 du per acre) in the area of analysis in year t. ☐

5. Compute the acreage of land within 5,000 feet of any interceptor sewer line in the area of analysis in year t. ☐

BOMA DATA

1. Enter the vacancy rate for office buildings in the area of analysis in 1960. ☐ %

CONSUMER PRICE INDEX

1. Enter the consumer price index for the community served and year t. ☐

2. Enter the consumer price index for the community served and the year of federal funding. ☐

EXHIBIT 3-1

## SURVEY FORM OF LAND USE FROM AERIAL PHOTOGRAPHS

Community Served by Major Project ________________________________________

Resources:
- Aerial photographs for area of analysis (1970)
- USGS 1:24,000 topographic map with boundaries of area of analysis shown (1970
- Land use map of area (1970) - included with some project materials, otherwise ask regional planning agency if available
- USGS 1:250,000 land use/land cover map - included with some project materials
- American Hospital Association Guide, 1973
- Interview with regional planning agency.

1. Check scale of photograph by measuring known distance on USGS topographic map and photograph. Indicate scale: 1 inch = _______ feet

2. Indicate year of coverage: _______________

3. For the scale, determine the number of grid squares equivalent to the following areas:

   1,000 $ft^2$ = __________________

   50,000 $ft^2$ = __________________

   100,000 $ft^2$ = __________________

4. For the scale, determine the reference shadow length of a known 1-story structure = _______ inches

5. For each structure, record the number of grid squares of roof area and the number of stories (estimated from shadow length) on the worksheet in the appropriate land use category. Apportion unknown uses equally among probable categories as you are working and, separately, keep track of the amount you apportioned in the miscellaneous column on the worksheet. Note, you should ignore buildings in the following use categories:

   - Residential
   - Recreational - Parks
   - Transportation - Utilities
   - Government - Public Assembly
   - Universities - Trade Schools

6. Compute total building floor area in 1,000 $ft^2$ for each land use category and record on summary sheet.

## EXHIBIT 3-1

| 1,000 $Ft^2$ Floor Area | Category | LUC* | Definition |
|---|---|---|---|
| | Commercial | 52-59 | Retail trade |
| | | 62 | Personal services |
| | <50 K | 64 | Repair services |
| | | 66 | Contract construction services |
| | 50-100 K | | |
| | >100 K | | |
| | Office-Professional | 61 | Finance, insurance, real estate |
| | | 63 (-637) | Business services |
| | | 65 | Professional services (includes hospotals) |
| | | 692, 699 | Welfare - charitable services |
| | Manufacturing | 2, 3 | Manufacturing |
| | Wholesale-Warehousing | 51 | Wholesale trade |
| | | 637 | Warehousing and storage services |
| | Public and Private Schools | 681 | Nursery, primary and secondary schools |
| | Other | 15 | Hotels-motels |
| | | 691 | Churches |
| | | 71, 79 | Cultural facilities |

* Standard Land Use Codes, U.S. Urban Renewal Adm. and Bureau of Public Roads, 1965.

EXHIBIT 3-1

| Commercial | | | Office-Professional | Manufacturing | Wholesale-Warehousing | Public and Private Schools | Other |
|---|---|---|---|---|---|---|---|
| <50 K | 50-100 K | >100 K | | | | | |
| | | | | | | | |

## EXHIBIT 3-1

## TRANSPORTATION DATA COLLECTION FORM

Data is collected for only 10 of the 40 case studies.

For the area of analysis circa 1970 determine if zonal transportation studies have been done. If so, obtain a map showing the traffic zones in the area of analysis or mark those on your USGS map. If zonal data are not available, gather the following data by the smallest geographical unit available and obtain a map showing the boundaries of such units.

Collect data for the following variables. For each variable, record the values by zone on separate sheets of paper or Xerox the data directly from reports and attach. No structured data form has been provided since the number of zones may vary from 10 to 100 or more.

1. Pattern of Land Use

   a. Land use (acres, $ft^2$) by zone (tabular or map) for the following categories: residential, commercial, industrial, wholesale/ warehousing, schools, office, recreation.

2. Social-Economic Characteristics

   a. Total population by zone
   b. Total employment by zone
   c. Employment by occupation by zone
   d. Dwelling units per zone
   e. Median income by zone
   f. Stratified income by zone
   g. Automobile ownership by zone (registration and fleet mix - car, truck, bus)

3. Transportation Facilities

   a. Map of transportation network (local, collection, arterial, expressway)
   b. Traffic volumes (ADT ) and design capacity for each major roadway
   c. Average trip length by purpose
   d. Number of interzonal trips by purpose per day (home, work, shopping, social-recreation, school, personal business)
   e. Average vehicle speeds for each major roadway
   f. Zonal trip generation rates by land use activity (same categories as 1a).
   g. Zonal VMT

Alt. Sources:

| | |
|---|---|
| 3a, b, and e | County or city traffic engineer |
| 3g | State department of motor vehicles |
| 1a, 2a-f | County or city planning departments |

## 3. Field Surveys

### a. Data Sources

The survey forms in the previous section were used by the Walden staff in personal visits to each major project site to gather data from local/regional planning agencies, municipal department of public works/sewer commissions/chamber of commerce, regional river basin commissions/water resources councils, and state transportation (3C) planning agencies. The data collected on-site consisted principally of information on the major project, land use in the area of analysis (from aerial photographs, published reports, and special agency unpublished reports), various regional parameters, watershed boundaries, and VMT data for the GEMLUP VMT model validation task. We collected field data for 36 of the 40 cases; the Project Officer was responsible for the remaining four cases. Transportation data were collected for 11 of the 40 case studies.

One key data item needed to define the area of analysis for each case study was a watershed map for the area. There is no single source for watershed maps nationwide. Thus, the combined sources listed below provided adequate data:

- The U.S. Geological Survey (USGS) has published a series of "State Hydrologic Unit Maps" showing the drainage basin boundaries, rivers and streams of all river basins exceeding 700 square miles in area. These maps have been prepared at a scale of 1:500,000 and were available for 24 eastern states* and the District of Columbia. Although at too small a scale to always identify the drainage basins for case study major projects directly, these maps provided guidance in this effort.
- Although the USGS does not print watershed boundaries on their quadrangle maps, these were deduced from elevations. Also, most USGS district offices had local drainage area maps which could be reviewed on-site, but not purchased.
- In some states, Regional River Basin Commissions, Water Resources Councils or Water Districts had done watershed mapping as part of their river basin studies.

---

*Alabama, Florida, Georgia, Illinois, Indiana, Maine, Maryland, Delaware, Massachusetts, Rhode Island, Connecticut, Minnesota, Michigan, New Hampshire, Vermont, New Jersey, New York, North Carolina, Ohio, Pennsylvania, South Carolina, Virginia, West Virginia, Wisconsin.

Supplementing the aerial photographs used to estimate land use in the areas of analysis were USGS land use/land cover maps at a scale of 1:250,000 for certain portions of the country. Overlay maps with hydrological units and/or census tract boundaries were also obtained. Each of these maps required a USGS topographical map for reference purposes. Although these land use maps could not be used as a direct source of land use data due to the small map scale and the fact that the roof area of urban structures is not shown, with overlays of hydrological unit information, they provided guidance in identifying the type of urban land use for parcels in the area of analysis of a major project. In addition, an overlay of census tract boundaries on land use helped provide an inventory of all census tracts contained in the area of analysis of a major project. This information was needed for the census data collection effort.

The land use/land cover maps published by the USGS are based on remote sensing data, viz. LANDSAT satellite images and high-altitude, color-infrared aerial photographs. Using these sources, data are presented at the two levels of detail shown in Table 3-5. It is noteworthy that the USGS classification system is resource oriented (i.e., it emphasizes land cover), while, by contrast, the Standard Land Use Coding (LUC) Manual [109] is people oriented (i.e., it emphasizes land use). Still, there is a degree of compatbility between the first level categories of the Standard Land Use Coding Manual and the second level categories under "Urban or Built-up Land" of the USGS classification system. Since the GEMLUP endogenous land use variables are defined in terms of LUC categories, the correspondence between the USGS and GEMLUP classification systems can be determined. This comparison is given in Table 3-6 which shows that USGS categories 11, 13, and 14 can indicate the dominance of one specific GEMLUP land use in a given area. The other USGS categories, although less informative because they indicate the dominance of any one of several possible GEMLUP land use types, still provided some guidance in the land use identification process.

TABLE 3-5

US GEOLOGICAL SURVEY LAND USE AND LAND COVER CLASSIFICATION SYSTEM

| Code | Level I Description | Code | Level II Description |
|---|---|---|---|
| 1 | Urban or Build-up Land | 11 | Residential |
| | | 12 | Commercial and Services |
| | | 13 | Industrial |
| | | 14 | Transportation, Communications, and Utilities |
| | | 15 | Industrial and Commercial Complexes |
| | | 16 | Mixed Urban or Build-up Land |
| | | 17 | Other Urban or Build-up Land |
| 2 | Agricultural Land | 21 | Cropland and Pasture |
| | | 22 | Orchards, Groves, Vineyards, Nurseries, and Ornamental Horticultural Areas |
| | | 23 | Confined Feeding Operations |
| | | 24 | Other Agricultural Land |
| 3 | Rangeland | 31 | Herbaceous Rangeland |
| | | 32 | Shrub and Brush Rangeland |
| | | 33 | Mixed Rangeland |
| 4 | Forest Land | 41 | Deciduous Forest Land |
| | | 42 | Evergreen Forest Land |
| | | 43 | Mixed Forest Land |
| 5 | Water | 51 | Streams and Canals |
| | | 52 | Lakes |
| | | 53 | Reservoirs |
| | | 54 | Bays and Estuaries |
| 6 | Wetland | 61 | Forested Wetland |
| | | 62 | Nonforest Wetland |
| 7 | Barren Land | 71 | Dry Salt Flats |
| | | 72 | Beaches |
| | | 73 | Sandy Areas other than Beaches |
| | | 74 | Bare Exposed Rock |
| | | 75 | Strip Mines, Quarries, and Gravel Pits |
| | | 76 | Transitional Areas |
| | | 77 | Mixed Barren Land |
| 8 | Tundra | 81 | Shrub and Brush Tundra |
| | | 82 | Herbaceous Tundra |
| | | 83 | Bare Ground Tundra |
| | | 84 | Wet Tundra |
| | | 85 | Mixed Tundra |
| 9 | Perennial Snow or Ice | 91 | Perennial Snowfields |
| | | 92 | Glaciers |

Source: [124].

## TABLE 3-6

## COMPATIBILITY BETWEEN THE U.S. GEOLOGICAL SURVEY AND GEMLUP LAND USE CLASSIFICATION SYSTEMS

| U.S. Geological Survey Code | U.S. Geological Survey Description | | GEMLUP Description |
|---|---|---|---|
| 11 | Residential | = | Residential |
| 12*† | Commercial and services | = | Commercial<br>Office-Professional Services<br>Wholesale-Warehousing<br>Educational<br>Active, Outdoor Recreation<br>Other Urban Land Uses |
| 13 | Industrial | = | Manufacturing |
| 14** | Transportation, communications, and utilities | = | Non-expressway highway |
| 15* | Industrial and commercial complexes | = | Commercial<br>Office-Professional Services<br>Wholesale-Warehousing<br>Manufacturing |
| 16* | Mixed urban or built-up land | = | Residential<br>Commercial<br>Office-Professional Services<br>Wholesale-Warehousing<br>Manufacturing |
| 17† | Other urban or built-up land | = | Education<br>Active, Outdoor Recreation<br>Other Urban Land Use |

---

* USGS codes 12, 15 and 16 also include LUCs 67 (Governmental Services), 682 (Colleges) and 683 (Trade Schools), not in GEMLUP

** USGS code 14 includes all of LUC 4, not just 45 (Highway and street right-of-way)

† USGS codes 12 and 17 also include LUC 76 (Parks), not in GEMLUP

Source: [124]

### b. Test-Training Case

As a prelude to the actual data collection, a training-test case was undertaken involving the Project Officer, the Walden program manager, and all field personnel. The objectives of the training case were to:

- Collect data for one of the 40 case studies,
- Test the feasibility of collecting the requisite data,
- Verify the time estimates developed for the data collection, task,
- Provide training to the field personnel, and
- Isolate any potential problem areas for solution.

The actual data collection process was then initiated and required in excess of 20 man-weeks of effort to complete.

### c. Results

Early on in the data collection, at the Project Officer's direction, data collection for soil suitability (item #10 on the Regional Planning Agency Survey Form) was discontinued. Field staff observed that urban development in the period 1960-1970 occurred on land having a "severe" soil type classification without regard to suitability for urban development. Since such classifications did not really restrict urban development in the study period, and since soil restrictions are probably better in explaing why a sewage treatment plant is constructed rather than the development that follows, SOILS and ONLOT were not appropriate exogenous variables for the model. Thus, the data needed to represent these variables was not collected. Additional reasons for not collecting the data were difficulty in obtaining data for just the area of analysis, lack of U.S. Soil Conservation Service surveys in most regions and lack and comparability of soil classifications in different regions.

The transportation data collected for 11 of the 40 case studies was examined and the data available (and thus obtained) is summarized in Table 3-7. The data are slightly less consistent than shown since variable definitions often varied between case studies, in terms of categories and units.

## TABLE 3-7

## SUMMARY OF GEMLUP CASE STUDY TRANSPORTATION DATA

| Transportation Case Studies | Variables for Which Data Were Available and Collected (See Page 3-17) | | | | | | | | | | | | | | | | Zonal Map |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | | | | | | | 3 | | | | | | | | |
| | | a | b | c | d | e | f | g | a | b(1) | b(2) | c | d | e | f | g | |
| Lexington, MA[1] | | | | | | | | | ✓ | ✓ | ✓ | | ✓ | ✓ | | | |
| Bennington, VT[2] | ✓ | | | | ✓ | | | | ✓ | ✓ | | | ✓ | | | | ✓ |
| Richmond, VA | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓[6] | ✓ | | | | ✓ |
| Charlotte, NC | | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ | | | ✓[3] | | | | ✓ |
| Clearwater, FL | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓[5] | ✓ | ✓ |
| Roseville, MN | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | | ✓ |
| St. Louis, MO | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓[4] | | | ✓[4] | ✓[5] | | ✓ |
| Boulder, CO | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ | ✓ | | ✓ | ✓[5] | | | | ✓ |
| S. Phoenix, AZ | | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ | ✓ | | ✓ | ✓[5] | ✓ | ✓[8] | | ✓ |
| Vallejo, CA | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | | ✓ | | | | ✓ |
| Auburn, WA | ✓ | ✓ | ✓ | ✓ | ✓ | | | | ✓ | | | ✓[7] | ✓ | | ✓[8] | ✓ | ✓ |

[1] Substitution for Hudson, MA.

[2] Extra transportation case study.

[3] Work trips only.

[4] Major arterials only.

[5] For area of analysis as a whole.

[6] Units are in elapsed time, not distance.

[7] For all trips as a whole.

[8] Trip generation equations for area as a whole by trip purpose, not land use.

Missing data from U.S. Census publications caused four separate problems. First, the lack of data for encost eliminated ENERGY from the model. Second, the lack of data for copopl eliminated DENCHG. However, a similar variable named PERCHG was introduced as a replacement. Third, 9 of 40 values for wwemp were unavailable. Lack of complete data would have eliminated two important variable WWJOB and RELWW from the model*. Thus, estimates for these nine values were obtained from SMSA data as follows:

$$wwemp = rwwemp\ (acre/areal)$$

Finally, missing data was a problem in 1970 census tract housing data by family unit size (variables unit34, unit 5, and unit50). These missing data did not effect the path analysis, but reduced the sample size for the variables MF and SFATT used in the Phase IV disaggregation analysis of the predictive equations.

### 4. Create Computer Data File

This task involved error checking and loading of the GEMLUP-II data files onto our computer system. Keypunched cards of data were loaded onto disk files and a program was written and executed to transform these raw data into a file of the variables chosen for the initial path model. Next each variable was analyzed using the CONDESCRIPTIVE command of SPSS and the data statistics (e.g. variable means and extremes) examined for errors. Several errors in the coding forms and the transformation program were discovered. All errors detected in this quality assurance check were corrected. The final GEMLUP-II data file is reproduced in Appendix E and the output of the final CONDESCRIPTIVE run is given in Appendix F. In order to test the initial path model in a non-linear, multiplicative form, a second data file was created. This file is identical to the first except that the natural logarithm of every data value was computed and stored.

---

*Although SPSS has the capability to handle missing data in statistical analyses, TSP does not and TSP is eventually used to analyze each land use category.

# IV. PHASE III - CAUSAL ANALYSIS

The principal objective of the third phase of this study was to test the initial causal model using the set of case study data and the statistical techniques of path analysis. The approach involved determining which of the hypothesized causal relationships were significant and selecting the parameters for the final causal model. The final path analytic model for induced land use was also used to trace the direct and indirect effects that variables have on each other and summarize the total net effects of such relationships. The final task was to validate the GEMLUP-I VMT model [19] using data collected from 11 of the case study major projects.

## A. PATH ANALYSIS

The objective of this task was to test the initial causal model developed in Phase I with the statistical techniques of path analysis. A full discussion of the concepts, assumptions, and problems associated with path analysis and its applicability to the current study is contained in Appendices A and B. Two statistical techniques were used to verify the hypothesized path models of induced land use and to determine model parameters: two-stage least squares and ordinary least squares multiple regression. The first technique was used to solve for path coefficients in the system of non-recursive equations connected by feedback loops. To solve for path coefficients in the other model equations that are not interconnected, ordinary least squares regression techniques was employed. All variables were standardized (to a mean of 0 and standard deviation of 1) prior to the path analysis. Thus, the analysis yielded standardized path coefficients ($\beta$ weights) as the model parameters. This approach was selected since standardized path coefficients ($\beta$ weights) provide an effective way of comparing the relative importance of various causal variables, independent of different units and scales. Thus, path coefficients can be used to judge the significance of model paths in theory trimming, the second task of Phase III.

Two different computer software packages were employed for the statistical analysis in this task. The first of these is the Time Series Processor (TSP) developed at Harvard University and the Massachusetts Institute of Technology (MIT) [125]. TSP is a programming language oriented towards the statistical analysis of time series, with specific applications to econometric research. TSP was used for all two stage least squares analysis in the GEMLUP study. The Statistical Package for the Social Sciences (SPSS) [92] was also used in the path analysis due to the comprehensive system of programs it provides for statistical data analysis.

### 1. Preselection of Exogenous Variables

The initial path analytic model consists of 9 structural equations, of which 7 are simultaneous and 2 are recursive. The complete model contains 9 endogenous and 32 exogenous variables. An additional 34 exogenous variables serve as alternatives to those specified in the initial model. Before testing the initial model, it was necessary to choose between the prime and alternative independent variables in each equation. The selection was based on the results of a zero-order correlation analysis for all model variables (see Appendix G). The selection process used was as follows:

- First, eliminate those independent variables whose correlation with the dependent variable is opposite in sign to that hypothesized. If all signs were incorrect, the prime variable was retained.
- Then, compare correlation values of the remaining independent variables and select the variable with the highest absolute value. If two or more variables have similar correlation values, the one thought to have the strongest causal relationship was chosen.

The effects of this task were to substitute many alternative variables for the prime independent variables in the initial model. A complete discussion of variables in model equations is given in section IV.A.5.c.

### 2. Identification

In order to obtain consistent estimates of the model coefficients in the two-stage least squares analysis, it was necessary that each equation satisfy the condition of identification which ensures an adequate number of instrumental variables. This condition is stated as:

$$m_0 \geq q-1$$

where:

$m_0$ = the number of instrumental variables, i.e., the number of exogenous variables which do not appear as independent variables in an equation

q = the number of independent, endogenous variables in an equation

Due to the substitutions of independent variables, the total set of instrumental variables in the simultaneous equation block increased to 32 (see Table 4-1a). This increase reduced the available degrees of freedom for model testing of any one simultaneous equation to 7 (=40-32-1). This still exceeds the minimum of 6 degrees of freedom thought to be necessary to obtain meaningful results in the initial path analysis. As variables were trimmed, the set of instrumental variables eventually was reduced to 12 (see Table 4-1b).

Application of the identification condition to each initial simultaneous equation indicates all are properly identified (see Table 4-2). The identification condition was reapplied at each step of the theory trimming task (see Section IV.A.5).

### 3. Multicollinearity

To identify potential problems of multicollinearity, the zero-order correlations between independent variables in each simultaneous equation were examined. Values of $|r| \geq 0.80$ usually indicate a problem.

## TABLE 4-1a

## INSTRUMENTAL VARIABLES USED IN THE INITIAL TWO-STAGE LEAST SQUARES ANALYSIS

CAPAC1
LENGTH
PHASE
TLIMIT
PERCHG
PERCAP
POOR
NONHH
LAND
STAY
AREA
RZONED
DRIVE
JOBS
UNIV
TRANS
OZONED
TIME
POPDIF
POPDEN
VACOFF
PEAK
MANJOB
RRMILE
ACCESS
IZONED
LIMITS
JOBCHG
EMPOP
DISCBD
VALUE
KIDS

## TABLE 4-1b

## INSTRUMENTAL VARIABLES USED IN THE FINAL TWO-STAGE LEAST SQUARES ANALYSIS

RECAP1
NONHH
POPDEN
PERCHG
LAND
RZONED
DRIVE
ACCESS
CZONED
JOBS
INCOME
OZONED

## TABLE 4-2

## CONFIRMATION OF SUFFICIENT IDENTIFICATION IN THE INITIAL TWO-STAGE LEAST SQUARES ANALYSIS

| Dependent Variable | Number of Instrumental Variables ($m_o$) | Number of Independent Endogenous Variables (q) | Sufficiency Condition $m_o \geq q-1$ |
|---|---|---|---|
| RES | 18 | 4 | $18 \geq 3$ |
| COMM | 25 | 3 | $25 \geq 2$ |
| OFFICE | 24 | 4 | $24 \geq 3$ |
| MANF | 21 | 1 | $21 \geq 0$ |
| HIWAYS | 26 | 4 | $26 \geq 3$ |
| EDUC | 25 | 1 | $25 \geq 0$ |
| REC | 27 | 2 | $27 \geq 1$ |

Examination of the correlation output shown in Appendix G revealed that most values of $|r|$ are less than 0.30, with two instances where it approached or exceeded 0.80. The first of these occurred between MANF and COMM ($r = 0.90$). This pair of independent variables appears in three equations (OFFICE, HIWAYS and WHOLE). The second case of multicollinearity occurred between RES and EDUC ($r=0.78$), both independent variables in the REC equation.

There are two solutions to a problem of multicollinearity. The first is to form a single composite variable which represents both variables (e.g., using factor analysis), while the second is to simply eliminate one of the pair. The first approach, if used, would not allow a complete statement of causality since composite variables are often hard to interpret physically. Thus, the second approach was used, i.e., one of the two intercorrelated variables was removed, prior to the first path analysis.

In the OFFICE equation, the following simple correlations are relevant:

$$r_{\text{OFFICE, MANF}} = 0.69 \qquad r_{\text{RES, MANF}} = 0.51$$
$$r_{\text{OFFICE, COMM}} = 0.71 \qquad r_{\text{RES, COMM}} = 0.50$$
$$r_{\text{HIWAYS, MANF}} = 0.30$$
$$r_{\text{HIWAYS, COMM}} = 0.39$$

These statistics reveal that while COMM is more highly associated with OFFICE than MANF; COMM is also associated with slightly more redundancy* in the equation. Causally, MANF is secondary to RES as a market generator for Office-Professional services, while COMM is the primary locational factor. Thus, MANF was eliminated from the equation.

In the HIWAYS equation, the above listed correlations are relevant. These statistics reveal that COMM is more highly associated with

---

*As exhibited by correlation with other independent variables, such as RES and HIWAYS in this example.

HIWAYS than MANF, and that both are associated with about equal redundancy in the equation. Causally, both variables represent generators of motor vehicle traffic, with the relationship between COMM and HIWAYS thought to be slightly stronger. Thus, MANF was eliminated from the equation.

In the WHOLE equation, the following correlations are relevant:

$r_{\text{WHOLE, MANF}} = 0.42$ $r_{\text{HIWAYS, COMM}} = 0.39$

$r_{\text{WHOLE, COMM}} = 0.25$ $r_{\text{HIWAYS, MANF}} = 0.30$

These statistics reveal that MANF is more highly associated with WHOLE than COMM, and that COMM is associated with more redundancy in the equation. Causally, both represent a portion of the same market cycle in which wholesale-warehousing development occurs. Thus, COMM was eliminated from the equation.

In the REC equation, the following correlations are relevant:

$r_{\text{REC, RES}} = 0.42$ $r_{\text{RES, POOR}} = 0.26$

$r_{\text{REC, EDUC}} = 0.34$ $r_{\text{EDUC, POOR}} = 0.53$

These statics indicate that RES is more highly associated with REC than EDUC, and that EDUC is associated with more redundancy in the equation. Causally, EDUC is secondary to RES as a generator of demand for recreational space. Thus, EDUC was eliminated from the equation.

Testing for multicollinearity was also performed after the initial path analysis by identifying any unusually large (in absolute value) beta weights and standard errors of the model coefficients.

### 4. Suppressor Variables

Suppressor variable problems generally occur in simultaneous systems when an independent variable is included which is causally unrelated to the dependent variable of the equation in which it appears. To identify potential problems, the zero-order correlations between dependent and independent variables were examined. Correlations not significant at the 20% level (i.e., $|r| < 0.14$) were assumed to indicate a potential problem. A list of the potential problem variables is given in Table 4-3. It is not desirable to trim these variables from the model using the simple correlations as the only guide; intercorrelations may be masking significant relationships between variables. Thus, at this point, it was sufficient to note where problems may occur so that they could be addressed in the theory trimming of the model.

### 5. Theory Trimming

The objectives of this task were to examine the statistical results of the initial path analysis to determine which paths are insignificant and redundant, and to trim these paths from the model.

#### a. Approach

Our approach to the execution of this task was as follows. First, a set of statistical criteria were developed to determine which paths should be trimmed from the structural equations. Next these criteria were applied to the results of the initial path analysis and the model refined. Additional path analyses were then performed until a final path model was determined. Thus, the trimming procedure was an iterative process. For example, at each step the remaining path coefficients were examined to see how the deletion of a path affected the ability of the model to reproduce the original observed correlations; this is particularly critical in two-stage least squares.

In deciding what paths to drop, we did not choose one fixed significance level (e.g., five percent) against which to compare the model t and F statistics. In the two stage least squares analysis, the values calculated for a t-statistic can only approximately* be interpreted as being distributed as t, due to the first stage estimation of all feedback loop variables. Thus, it was necessary to use a more general rule in

---

*Asymptotically

## TABLE 4-3

## INDEPENDENT VARIABLES NOT SIGNIFICANTLY CORRELATED WITH DEPENDENT VARIABLES IN THE SIMULTANEOUS BLOCK OF EQUATIONS

| Dependent Variable | Independent Variables |
|---|---|
| RES | STAY, LENGTH, PHASE |
| COMM | POPDIF, TIME, UNIV, OZONED |
| OFFICE | TIME, JOBCHG, VACOFF |
| MANF | LAND, LIMITS, LENGTH, TIME |
| HIWAYS | ACCESS, DRIVE, DISCBD, UNIV, EMPOP, TIME |
| EDUC | PERCHG, NONHH, TIME |
| REC | UNIV |
| WHOLE | LAND, COMCHG, TIME |
| OTHER | TRANS, INCOME, EMPOP, UNIV, TIME |

interpreting these statistics. In ordinary least squares analysis, an accepted guideline is that if t or F* is greater than 1.0, the variable being tested raises the adjusted $R^2$ of the equation, i.e., it explains a significant additional amount of the variance of the dependent variable. Conversely, a t or F less than 1.0 indicates that the variable in question lowers the adjusted $R^2$ of the equation.

The adjusted $R^2$ is defined as:

$$R^2_a = R^2 - \left(\frac{k-1}{N-k}\right)(1-R^2)$$

where:

k = number of independent variables

N = number of data samples.

The conventional $R^2$ statistic can yield deceptive results when the significance of independent variables is in question. For example, simply adding a variable to any regression equation, whether it is at all correlated with the dependent variable or not, will raise the $R^2$ and indicate additional variance has been explained. The $R^2_a$ gives a more conservative, unbiased estimate of the amount of variance explained in the dependent variable through the regression equation.

The path coefficient ($\beta$) indicates the expected standard deviation change in the dependent variable given a one standard deviation change in a particular independent variable, all other effects held constant. Comparing the $\beta$'s for different variables allows one to evaluate the relative importance of the various independent variables because of the standardization. In general, $\beta < 0.1$ indicates a variable is not contributing importantly to the regression equation.

In trimming the path analysis model then, the following tests were applied:

---

*$F(1, N-k-1) = t^2(N-k-1)$

- A test to ensure statistical significance, $|t|$ of $F \geq 1.0$.

- A test to ensure independent variables make a significant contribution, $\beta \geq 0.1$.

In addition to the above, the following criteria were also applied:

- A test to ensure a maximum value for the model parameters, $\beta \leq 1.0$, indicating no redundant paths through excessive independent variable correlation.

- A test to ensure the sign of the path coefficient ($\beta$) is correct and not counter to the original model hypothesis.

- A test to ensure a path being trimmed does not totally eliminate an exogenous or instrumental variable from the model which is deemed *a priori* to be of substantive causal importance.

### b. Results

The results of the final path analysis are shown in the form of the final causal path diagram in Figure 4-1. The path coefficients ($\beta$) are shown on each path and the $R^2$ statistic for each equation displayed in the box of the associated dependent variable. The results indicate that the final model explains the more than half of variance in case study data with $R^2$ values ranging from 0.27 to 0.82 and averaging 0.54. The residuals of the final regressions do not exhibit any trends or patterns, indicating the remaining unexplained variance ($1-R^2$) is not due to poor specification of the model, but rather due to wide variance in the case study data (i.e. the problem of trying to develop one generalized model for a broad range of situations). Several of the equations do have one or two very large residuals which account for 30-50% of the unexplained variance by themselves (see Table 4-4). An examination of these data points did not uncover any errors in input data, but it did reveal that most large residuals correspond to the highest or second-highest values for the dependent variable and that these are all more than three standard deviations away from the mean, i.e. they are extreme values. An examination of the distributions

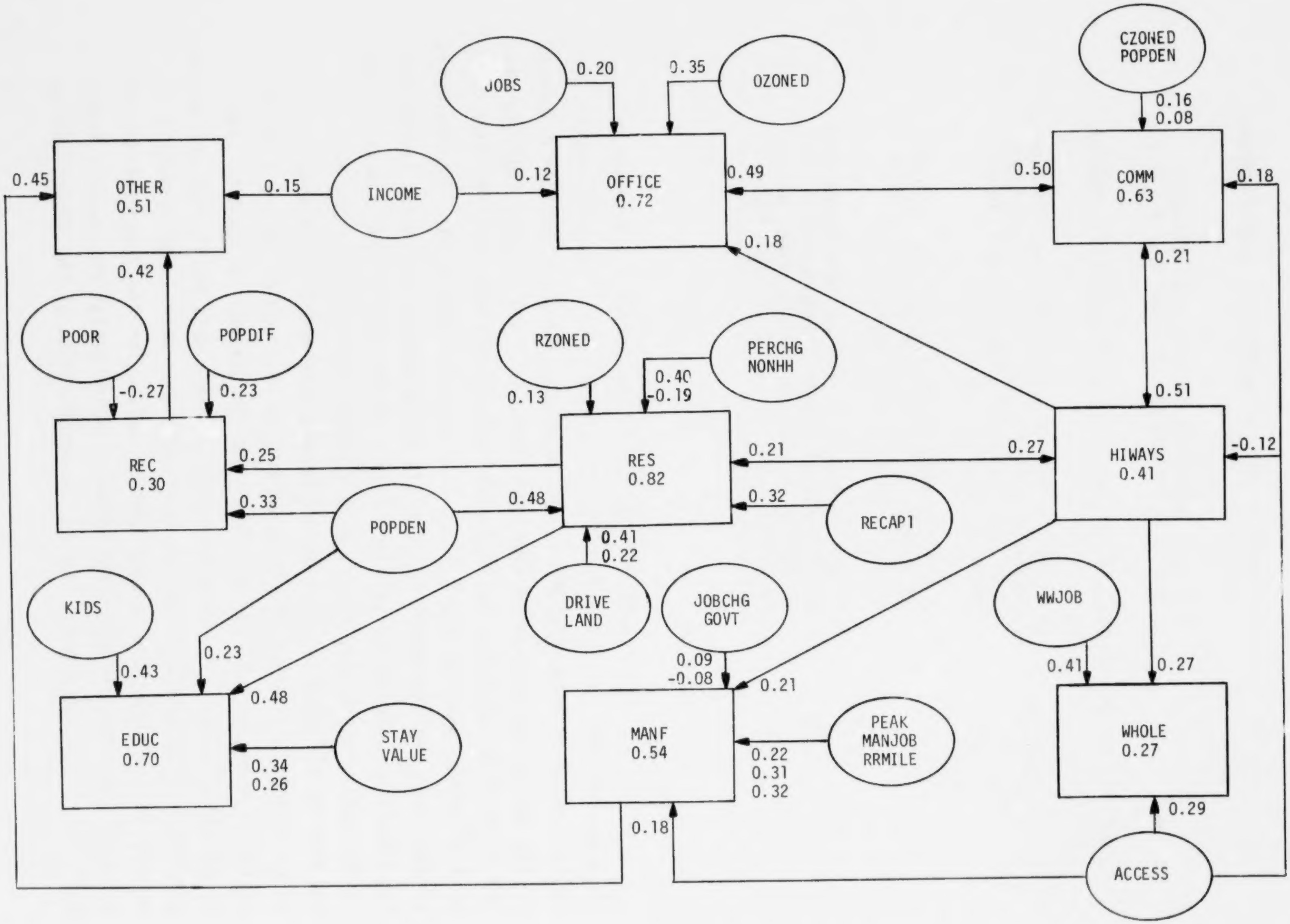

JOBS
0.20
0.35
OZONED
CZONED
POPDEN
0.16
0.08
0.45
OTHER
0.51
0.15
INCOME
0.12
OFFICE
0.72
0.49
0.50
COMM
0.63
0.18
0.18
0.21
0.42
POOR
POPDIF
RZONED
PERCHG
NONHH
-0.27
0.23
0.13
0.40
-0.19
0.51
REC
0.30
0.25
RES
0.82
0.21
0.27
HIWAYS
0.41
-0.12
0.33
POPDEN
0.48
0.32
RECAP1
0.41
0.22
DRIVE
LAND
JOBCHG
GOVT
KIDS
WWJOB
0.23
0.43
0.09
-0.08
0.21
0.41
0.27
EDUC
0.70
0.48
STAY
VALUE
0.34
0.26
MANF
0.54
PEAK
MANJOB
RRMILE
0.22
0.31
0.32
WHOLE
0.27
0.18
0.29
ACCESS

FIGURE 4-1

FINAL PATH ANALYTIC MODEL

## TABLE 4-4

## EXTREME VALUES WHICH ACCOUNTED FOR A SUBSTANTIAL PORTION OF UNEXPLAINED VARIANCE IN THE PATH ANALYSIS AND PRODUCED LARGE RESIDUALS

| Dependent Variable | Case Study Number of Extreme Value | Percent of Unexplained Variance Caused by Extreme Value |
|---|---|---|
| COMM | 35 | 48% |
| MANF | 35 | 37% |
| | 18 | 5% |
| WHOLE | 7 | 28% |
| | 14 | 18% |
| HIWAYS | 18 | 22% |
| EDUC | 3 | 38% |
| REC | 18 | 24% |
| OTHER | 18 | 39% |

for the dependent variables involved (all but RES and OFFICE) showed that they were skewed by these extreme values. Thus, there is a statistical basis for excluding these points from the model. Such exclusion would, however, restrict the model's applicability to a much narrower range of circumstances. Since it is the objective of this study to develop a generalized model, it was decided not to exclude the extreme value data. The complete statistical output of the final path analysis, including residual plots, is given in Appendix H.

In the course of trimming individual equations, only one instance of multicollinearity (i.e., $|\beta| \geq 1.0$) was observed. This involved the independent variables POPDEN and OFFJOB which appeared together in the COMM equation in the third path analysis. The intercorrelation of these two variables was strong (R = 0.93) and caused instabilities in the COMM equation. Of the two, POPDEN was retained because of its stronger causal and statistical relationship with COMM.

Of the potential suppressor variables uncovered through examination of the zero-order correlations, only five were not eliminated during the normal course of theory trimming. To test whether these five variables might be causing errors in the coefficients of other model variables, the equations in which they appeared were re-estimated, excluding them, to see if such changes did occur. In the case of three variables, STAY in the RES equation, POPDIF in the COMM equation and LENGTH in the MANF equation, very large changes did occur and so these suppressor variables were deleted (as discussed below). The exclusion of the other two variables, ACCESS in the HIWAYS equation and INCOME in the OTHER equation, produced changes in other model coefficients of 9% or less. Thus, it was concluded that a suppressor variable problem did not exist with these two and so they were retained.

### c. Discussion of Individual Equations

The actual path analysis was carried out using variables defined in English units. The path regression coefficients shown in the printout in Appendix H reflect this fact. The path coefficients ($\beta$) shown in Figure 4-1 are independent of the units chosen. The summaries of individual equations in this section give the path regression coefficient for each final model equation in both English and metric units.* The conversion to metric involved only the dimension of distance. The following conversion factors were used:

| | | |
|---|---|---|
| $\text{mile}^2$ | → $\text{km}^2$ | , multiply by 2.589 |
| $1{,}000\ \text{ft}^2$ | → $\text{m}^2$ | , multiply by 92.90 |
| acres | → $\text{m}^2$ | , multiply by 4,047 |
| miles | → km | , multiply by 1.609 |
| $\text{acre}^{-1}$ | → $\text{m}^{-2}$ | , divide by 4,047 |
| $\text{mile}^{-1}$ | → $\text{km}^{-1}$ | , divide by 1.609 |

In addition to equation coefficients, the following discussions also list the R, $R^2$ and coefficient of variation** for each final equation. The latter statistic is a measure of the precision of each model equation.

(1) RES Equation. In the first path analysis, REC, LENGTH, POOR and AREA were trimmed (low $\beta$,t statistics or incorrect sign). The sign for OFFICE was negative, counter to the original hypothesis. OFFICE was kept however because the statistical relationship with RES is very strong and can be explained if it is assumed that when a large amount of office development occurs, it sets the character of the area for further office/industrial development and not housing. The sign for LAND was positive, counter to the original hypothesis. LAND was kept though because the statistical relationship is strong and it is plausible that vacant developable land, which is on the edge of development, will have already risen in value in comparison to other parcels of land. Thus, residential development will more likely occur in areas where land prices are higher. The variable VACANT was added to replace AREA, representing the important causal factor

---

*A list of model variables giving both English and metric units is given in the following page.

**Standard error of the regression as a percentage of the dependent variable mean.

## LIST OF CAUSAL MODEL VARIABLES AND CORRESPONDING ENGLISH AND METRIC UNITS

| Variable Name | English Units | Metric Units |
|---|---|---|
| RES | Dwelling units per $10^4$ acres of land | Dwelling units per $10^4$ $m^2$ of land |
| COMM | $10^3$ $ft^2$ floor area per $10^4$ acres | $m^2$ floor area per $10^4$ $m^2$ of land |
| OFFICE | $10^3$ $ft^2$ floor area per $10^4$ acres | $m^2$ floor area per $10^4$ $m^2$ of land |
| MANF | $10^3$ $ft^2$ floor area per $10^4$ acres | $m^2$ floor area per $10^4$ $m^2$ of land |
| HIWAYS | Lane miles per $10^4$ acres | Lane km per $10^4$ $m^2$ of land |
| EDUC | $10^3$ $ft^2$ floor area per $10^4$ acres | $m^2$ floor area per $10^4$ $m^2$ of land |
| REC | Acres recreational land per $10^4$ acres | $m^2$ recreational land per $10^4$ $m^2$ of land |
| WHOLE | $10^3$ $ft^2$ floor area per $10^4$ acres | $m^2$ floor area per $10^4$ $m^2$ of land |
| OTHER | $10^3$ $ft^2$ floor area per $10^4$ acres | $m^2$ floor area per $10^4$ $m^2$ of land |
| RECAP1 | Percent** collection sys reserve cap. | * |
| NONHH | People per $mile^2$ of land | People per $km^2$ of land |
| POPDEN | People per $mile^2$ of land | People per $km^2$ of land |
| PERCHG | Percent+ change in county pop. | * |
| LAND | Unitless (land cost/regional income) | * |
| RZONED | Percent+ of area zoned residential | * |
| DRIVE | 100s of people per $mile^2$ of land | 100s of people per $km^2$ of land |
| ACCESS | Interchanges per $mile^2$ of land | Interchanges per $km^2$ of land |
| CZONED | Percent+ of area zoned commercial | * |
| JOBS | People per $mile^2$ of land | People per $km^2$ of land |
| INCOME | Unitless (tract family inc./county fam. inc.) | * |
| OZONED | Percent+ of area zoned office | * |
| RRMILE | Miles of track per $mile^2$ of land | km of track per $km^2$ of land |
| MANJOB | People per $mile^2$ of land | People per $km^2$ of land |
| PEAK | Million gallons per day (mgd) | * |
| JOBCHG | People per $mile^2$ of land | People per $km^2$ of land |
| GOVT | Millions of dollars ($10^6$ \$) | * |
| KIDS | People per 100 dwelling units | * |
| STAY | Percent+ of families in same house | * |
| VALUE | \$ (median value of housing) | * |
| POOR | Percent+ of families with low income | * |
| POPDIF | Percent+ change in regional pop. | * |
| WWJOB | People per $mile^2$ of land | People per $km^2$ of land |

---

*Same as English units.

**A value of 10% is inserted as 10.

+A value of 10% is inserted as 0.1.

of land constraints. The trimming of REC made the EDUC and REC equations recursive in subsequent analyses.

In the second path analysis, CAPAC1, TLIMIT, and VACANT were trimmed (low $\beta$,t statistics or incorrect sign). The variables RECAP1 and LIMITS were added as replacements, respectively. Also, the variables JOBS and SERVED were added to represent the important causal factors of employment characteristics and major project sewer services.

In the third path analysis, LIMITS, PHASE and SERVED were trimmed (low $\beta$,t statistics). In the fourth path analysis, MANF, OFFICE, PERCAP and JOBS were trimmed (low $\beta$,t statistics). CAPCHG and UNEMP were added to replace the latter two. Also, STAY was trimmed because of suppressor variable problems, i.e. its simple correlation with RES is insignificant and its inclusion in the equation causes instabilities in other model coefficients (e.g. betas change by more than 100%). The trimming of MANF makes the MANF equation recursive.

In the fifth path analysis, CAPCHG and UNEMP were trimmed (incorrect sign and low $\beta$,t statistics) and the final equation became:

Unstandardized

English

$$\begin{aligned} \text{RES} = {} & 40.36\ \text{HIWAYS} + 12.89\ \text{RECAP1} - 7.244\ \text{NONHH} \\ & + 2.007\ \text{POPDEN} + 10{,}419\ \text{PERCHG} + 2416\ \text{LAND} \\ & + 2798\ \text{RZONED} + 892.8\ \text{DRIVE} - 5{,}444 \end{aligned}$$

Metric

$$\begin{aligned} \text{RES} = {} & 25.08\ \text{HIWAYS} + 0.003185\ \text{RECAP1} - 0.004634\ \text{NONHH} \\ & + 0.001284\ \text{POPDEN} + 2.574\ \text{PERCHG} + 0.5970\ \text{LAND} \\ & + 0.6914\ \text{RZONED} + 0.5712\ \text{DRIVE} - 1.345 \end{aligned}$$

Standardized

$$\begin{aligned} \text{RES} = {} & 0.21\ \text{HIWAYS} + 0.32\ \text{RECAP1} - 0.19\ \text{NONHH} \\ & + 0.48\ \text{POPDEN} + 0.40\ \text{PERCHG} + 0.22\ \text{LAND} \\ & + 0.13\ \text{RZONED} + 0.41\ \text{DRIVE} \end{aligned}$$

$R = 0.91$, $R^2 = 0.82$, Coefficient of Variation = 0.32

(2) COMM Equation. In the first path analysis, RES, PERCAP, JOBS, UNIV, TRANS and TIME were all trimmed (low $\beta$,t statistics and/or incorrect sign). Three new variables were selected from the list of alternatives to replace trimmed variables representing the important causal factors of income, accessibility and zoning. These variables are INCOME, ACCESS and CZONED.

In the second path analysis, INCOME was trimmed (incorrect sign and low $\beta$,t statistics) and replaced by CAPCHG. Also, POPDEN and OFFJOB were added to represent the important causal factors of population and employment characteristics.

In the third path analysis, CAPCHG was trimmed (incorrect sign). Also, POPDIF was trimmed because of suppressor variable problems, i.e. its simple correlation with COMM is insignificant and its addition to the equation causes instabilities in other model coefficients (e.g. betas change by 100% with values approaching 1.0). In addition, OFFJOB was trimmed because of a multicollinearity problem. The final equation became:

Unstandardized

English
$$\mathrm{COMM} = 2.097\ \mathrm{OFFICE} + 13.98\ \mathrm{HIWAYS} + 3{,}082\ \mathrm{ACCESS} + 8{,}249\ \mathrm{CZONED} + 0.1206\ \mathrm{POPDEN} - 649.5$$

Metric
$$\mathrm{COMM} = 2.097\ \mathrm{OFFICE} + 807.2\ \mathrm{HIWAYS} + 183.2\ \mathrm{ACCESS} + 189.4\ \mathrm{CZONED} + 0.007167\ \mathrm{POPDEN} - 14.91$$

Standardized

$$\mathrm{COMM} = 0.50\ \mathrm{OFFICE} + 0.21\ \mathrm{HIWAYS} + 0.18\ \mathrm{ACCESS} + 0.16\ \mathrm{CZONED} + 0.08\ \mathrm{POPDEN}$$

$R = 0.80$, $R^2 = 0.63$, Coefficient of Variation = 0.75

(3) OFFICE Equation. In the first path analysis, RES, TRANS, POPDEN, TIME, JOBCHG and VACOFF were all trimmed (low $\beta$,t statistics and/or incorrect sign). Three new variables, ACCESS, OFFCHG and VACANT were added to replace trimmed variables representing the important causal factors of accessibility, regional influences and land constraints.

In the second path analysis, PERCAP and VACANT were trimmed (low β,t statistics) and replaced by CAPCHG and GOVT.

In the third path analysis, CAPCHG, OFFCHG and GOVT were trimmed (low β,t statistics). INCOME was added as a replacement for CAPCHG.

In the fourth path analysis, ACCESS was trimmed (low β,t statistics) and the final equation became:

Unstandardized

English
$$\mathrm{OFFICE} = 0.1149\ \mathrm{COMM} + 2.904\ \mathrm{HIWAYS} + 0.1853\ \mathrm{JOBS} + 265.8\ \mathrm{INCOME} + 22{,}133\ \mathrm{OZONED} - 390.9$$

Metric
$$\mathrm{OFFICE} = 0.1149\ \mathrm{COMM} + 167.7\ \mathrm{HIWAYS} + 0.01101\ \mathrm{JOBS} + 6.102\ \mathrm{INCOME} + 508.1\ \mathrm{OZONED} - 8.973$$

Standardized

$$\mathrm{OFFICE} + 0.49\ \mathrm{COMM} + 0.18\ \mathrm{HIWAYS} + 0.20\ \mathrm{JOBS} + 0.12\ \mathrm{INCOME} + 0.35\ \mathrm{OZONED}$$

$R = 0.85$, $R^2 = 0.72$, Coefficient of Variation = 0.56

(4) MANF Equation. In the first path analysis, AREA, IZONED, LIMITS, LAND and JOBCHG were all trimmed (low β,t statistics). Two new variables GOVT and MANCHG were added to represent the important causal factors of tax costs and regional influences.

In the second path analysis, TIME and MANCHG were trimmed (low β,t statistics) and JOBCHG was added to replace the latter.

In the third path analysis, LENGTH was dropped because of suppressor variable problems (i.e. change in other betas of up to 100% result from its inclusion in the equation).

In the fourth path analysis, JOBCHG and GOVT had marginally low β and t statistics, but were not trimmed because of their causal importance. The final equation became:

Unstandardized

English

$$\begin{aligned}\text{MANF} = {} & 1{,}984\ \text{RRMILE} + 3.754\ \text{MANJOB} + 111.3\ \text{PEAK} \\ & + 16.38\ \text{HIWAYS} + 3{,}627\ \text{ACCESS} + 5.969\ \text{JOBCHG} \\ & - 2.201\ \text{GOVT} - 1{,}647\end{aligned}$$

Metric

$$\begin{aligned}\text{MANF} = {} & 73.28\ \text{RRMILE} + 0.2231\ \text{MANJOB} + 2.555\ \text{PEAK} \\ & + 945.7\ \text{HIWAYS} + 215.6\ \text{ACCESS} + 3.547\ \text{JOBCHG} \\ & - 0.05052\ \text{GOVT} - 37.81\end{aligned}$$

Standardized

$$\begin{aligned}\text{MANF} = {} & 0.32\ \text{RRMILE} + 0.31\ \text{MANJOB} + 0.22\ \text{PEAK} \\ & + 0.21\ \text{HIWAYS} + 0.18\ \text{ACCESS} + 0.09\ \text{JOBCHG} \\ & - 0.08\ \text{GOVT}\end{aligned}$$

$R = 0.74$, $R^2 = 0.55$, Coefficient of Variation = 1.00

(5) HIWAYS Equation. In the first path analysis, DISCBD, UNIV, EMPOP, DRIVE, and TIME were all trimmed (incorrect sign). The variable POPDEN was added to replace UNIV, representing the important causal factor of population characteristics.

In the second path analysis, OFFICE and POPDEN were trimmed (low β, t statistics) and DUACRE was added to replace the latter. Also, the variables JOBS and AIRPRT were added to replace EMPOP and DISCBD.

In the third path analysis, AIRPRT, JOBS and DUACRE were trimmed (low $\beta$,t statistics and/or incorrect sign). The final equation became:

Unstandardized

English $$\begin{aligned}\text{HIWAYS} = {} & 0.001452\ \text{RES} + 0.007517\ \text{COMM} \\ & - 29.75\ \text{ACCESS} + 19.54\end{aligned}$$

Metric $$\begin{aligned}\text{HIWAYS} = {} & 0.002336\ \text{RES} + 0.01209\ \text{COMM} \\ & - 0.03062\ \text{ACCESS} + 0.007769\end{aligned}$$

Standardized

$$\begin{aligned}\text{HIWAYS} = {} & 0.27\ \text{RES} + 0.51\ \text{COMM} \\ & - 0.12\ \text{ACCESS}\end{aligned}$$

$R = 0.64$, $R^2 = 0.41$, Coefficient of Variation = 0.58

(6) EDUC Equation. In the first path analysis, NONHH, RZONED and PERCHG were trimmed (low $\beta$,t statistics or incorrect sign). The sign of STAY was positive, counter to the original hypotheses. STAY was kept however because the causal relationship is strong and can be explained as a secondary effect (after RES) where less mobile residential areas (i.e. more stable areas) are more likely to encourage educational programs and hence building development. The variable EDUCHG was added to replace PERCHG, representing the important causal factor of regional influences. The variable KIDS was redefined as school age children per 100 households in 1960 to allow a stronger causal relationship.

In the second path analysis, EDUCHG was trimmed (incorrect sign and low $\beta$,t statistics). The variables HSECHG and POPDIF were added as replacements, and POPDEN was added to represent population characteristics.

In the third path analysis, HSECHG, TIME, and POPDIF were trimmed (incorrect sign and/or low $\beta$,t statistics). The final equation became:

Unstandardized

English
$$\mathrm{EDUC} = 0.03587\ \mathrm{RES} + 8.059\ \mathrm{KIDS} + 1447\ \mathrm{STAY} + 0.03636\ \mathrm{VALUE} + 0.07166\ \mathrm{POPDEN} - 1713$$

Metric
$$\mathrm{EDUC} = 3.332\ \mathrm{RES} + 0.1850\ \mathrm{KIDS} + 33.22\ \mathrm{STAY} + 0.0008347\ \mathrm{VALUE} + 0.004259\ \mathrm{POPDEN} - 39.32$$

Standardized

$$\mathrm{EDUC} = 0.48\ \mathrm{RES} + 0.43\ \mathrm{KIDS} + 0.34\ \mathrm{STAY} + 0.26\ \mathrm{VALUE} + 0.23\ \mathrm{POPDEN}$$

$R = 0.84$, $R^2 = 0.70$, Coefficient of Variation = 0.40

(7). REC Equation. In the first path analysis, AREA and RZONED were trimmed (incorrect sign). A new variable, VACANT, was added to represent the important causal factor of land constraints.

In the second path analysis, UNIV, PERCHG and VACANT were trimmed (incorrect sign and/or low $\beta$,t statistics). The variables POPDEN, POPDIF and LAND were added as replacements, respectively.

In the third path analysis, LAND was trimmed (incorrect sign and low $\beta$,t statistics). The final equation became:

Unstandardized

English
$$\mathrm{REC} = 0.007844\ \mathrm{RES} - 805.0\ \mathrm{POOR} + 0.04382\ \mathrm{POPDEN} + 229.2\ \mathrm{POPDIF} + 127$$

Metric
$$\mathrm{REC} = 31.74\ \mathrm{RES} - 805.0\ \mathrm{POOR} + 0.1134\ \mathrm{POPDEN} + 229.2\ \mathrm{POPDIF} + 127$$

Standardized

$$\mathrm{REC} = 0.25\ \mathrm{RES} - 0.27\ \mathrm{POOR} + 0.33\ \mathrm{POPDEN} + 0.23\ \mathrm{POPDIF}$$

$R = 0.55$, $R^2 = 0.30$, Coefficient of Variation = 0.81

The low $R^2$ is possibly due to the fact that large, active recreational areas are usually dependent on governmental decisions and the availability of appropriate natural resources. Neither of these effects could be quantified for the causal model.

(8) WHOLE Equation. In the first path analysis, MANF, LAND, POPDEN, AREA, IZONED, TIME and COMCHG were all trimmed (low $\beta$,t statistics). The variables VACANT and MANCHG were added to represent the important causal factors of land constraints and regional influences.

In the second path analysis, MANCHG and VACANT were trimmed (low $\beta$,t statistics). The variable POPDIF was added to represent regional influences.

In the third path analysis, POPDIF was trimmed (incorrect sign) and the final equation became:

Unstandardized

English WHOLE = 9.022 WWJOB + 1,422 ACCESS + 5.151 HIWAYS - 94.28

Metric WHOLE = 0.5362 WWJOB + 84.51 ACCESS + 297.4 HIWAYS - 2.164

Standardized

WHOLE = 0.41 WWJOB + 0.29 ACCESS + 0.27 HIWAYS

$R = 0.52$, $R^2 = 0.27$, Coefficient of Variation = 1.12

(9) OTHER Equation. In the first path analysis, all variable except REC, MANF and INCOME were trimmed (low $\beta$,t statistics and/or incorrect sign). The variables ACCESS, POPDEN and SERCHG were added to represent the important causal factors of accessibility, population characteristics and regional influences.

In the second path analysis, POPDEN, ACCESS and SERCHG were trimmed (low $\beta$,t statistics and incorrect sign). The variables INTDEN, DRIVE and RIDET were added to represent accessibility, and the variables POPDIF and JOBCHG were added to represent regional influences. JOBS was added to represent employment characteristics.

In the third path analysis, JOBCHG, DRIVE, POPDIF, INTDEN and JOBS were trimmed (incorrect sign and low $\beta$,t statistics). In addition, RIDET was trimmed (low $\beta$,t statistics). The final equation became:

Unstandardized

English OTHER = 1.407 REC + 0.1176 MANF + 430.1 INCOME - 527.0

Metric OTHER = 0.03230 REC + 0.1176 MANF + 9.873 INCOME - 12.10

Standardized

OTHER = 0.42 REC + 0.45 MANF + 0.15 INCOME

$R = 0.72$, $R^2 = 0.52$, Coefficient of Variation = 1.02

### 6. Comparison of Model Forms

The first alternative model form considered was a mixed rates and totals model (viz., the form used in GEMLUP-I) in which the endogenous land use variables are not normalized by case study area. The initial path analysis was rerun using the mixed rates/totals form, and the results compared with that from the initial path analysis of the pure rates model. Comparisons were made using a coefficient of variation statistic (the standard error as a percent of the dependent variable mean) and are shown in Table 4-5. The results indicate a lower coefficient of variation in six out. of nine equations using a pure rate model. Thus, the first alternative model form was rejected.

The second alternative form considered was a multiplicative model of the form:

$$Y_1 = a\, Y_2^b\, X_1^c\, X_2^d \;.....$$

where $Y_1$, $Y_2$ are endogenous variables, $X_1$ and $X_2$ are exogenous variables, and a,b,c and d are model parameters. The initial path analysis was rerun using the multiplicative form (by taking the natural logarithm of all variables), and the results compared with that from the initial path analysis of the additive model. Due to the non-linearities in the multiplicative form, it was inappropriate to compare standard statistical measures. Instead, the residuals for each form were examined to determine how many occurred outside the standardized bounds of -1.0 or +1.0, see Table 4-6. This comparison of residuals can be thought of as a goodness of fit test, similar to the

### TABLE 4-5

### COMPARISON OF COEFFICIENTS OF VARIATION FOR A PURE RATES AND A MIXED TOTALS-RATES MODEL

| Dependent Variable | Pure Rates Model | Mixed Totals-Rates Model |
|---|---|---|
| RES | 0.44 | 0.55 |
| COMM | 0.78 | 1.03 |
| OFFICE | 0.60 | 0.80 |
| MANF | 1.08 | 1.20 |
| HIWAYS | 0.64 | 0.58 |
| EDUC | 0.42 | 0.54 |
| REC | 0.84 | 1.10 |
| WHOLE | 1.13 | 1.02 |
| OTHER | 1.03 | 0.67 |

*Endogenous variables as totals, all exogenous variables as rates.

### TABLE 4-6
### COMPARISON OF THE NUMBER OF STANDARDIZED RESIDUALS GREATER THAN 1.0 IN ABSOLUTE VALUE FOR TWO MODEL FORMS

| DEPENDENT VARIABLE | ADDITIVE MODEL | | | MULTIPLICATIVE | | |
|---|---|---|---|---|---|---|
| | <-1.0 | >+1.0 | Total | <-1.0 | >+1.0 | Total |
| RES | 6 | 4 | 10 | 5 | 9 | 14 |
| COMM | 3 | 4 | 7 | 6 | 6 | 12 |
| OFFICE | 3 | 4 | 7 | 5 | 5 | 10 |
| MANF | 4 | 3 | 7 | 8 | 6 | 14 |
| HIWAYS | 6 | 5 | 11 | 5 | 6 | 11 |
| EDUC | 4 | 1 | 5 | 4 | 6 | 10 |
| REC | 4 | 6 | 10 | 7 | 7 | 14 |
| WHOLE | 3 | 4 | 7 | 3 | 1 | 4 |
| OTHER | 1 | 3 | 4 | 2 | 3 | 5 |

$R^2$ statistic. The results indicate a lower number of large residuals in eight out of nine equations using the additive form. In addition, some of the residuals were plotted versus their dependent variables to check for any patterns. No distinct trends were apparent, though the residuals from the additive model did appear to be more randomly distributed. Thus, the multiplicative form was rejected.

### B. COEFFICIENT STABILITY ANALYSIS

The usefulness of any path analytic model is dependent on the stability of the path estimates. In the current study, where only a relatively small data sample (40) was available for estimating model parameters, the question of stability was therefore important. The statistical technique chosen to investigate stability was jackknifing. This procedure involved 40 separate analyses of each model equation using only 39 data samples. For each analysis, a different sample from the total set of 40 was excluded. This approach provided 40 separate estimates of each of the model coefficients, from which mean and standard deviations, and one-way frequency distributions were prepared. Two separate indices of coefficient stability were then developed for interpreting the data. The jackknifing analysis was applied to all nine equations in the model. In the jackknifing analysis, only the path regression coefficients (b) were used. Path coefficients ($\beta$) do not provide a valid basis for comparisons of systems operating on different data populations since the variance of the independent variables is included in the path coefficients and variance can change dramatically from sample to sample.

The means and standard deviations computed from the set of 40 values for each model coefficient are compared on Table 4-7 with the coefficient values from the final path analysis. The results show the jackknifing results and final path analysis path regression coefficients to be practically identical in most cases, with the exception of the COMM equation where differences are apparent. The standard deviation of the jackknifing coefficients is, on the average, only 15 percent of the coefficient mean value, indicating low variance in the coefficient values and thus no significant instabilities in general.

## TABLE 4-7

## COMPARISON OF PATH REGRESSION COEFFICIENTS FROM THE FINAL PATH ANALYSIS AND THE JACKKNIFING STABILITY ANALYSIS

| Model Equation | Independent Variable | Path Regression Coefficients* | | |
|---|---|---|---|---|
| | | Final Path Analysis | Jackknifing Mean | Jackknifing Standard Deviation |
| RES | NONHH | -7.244 | -7.311 | 0.505 |
| | HIWAYS | 40.36 | 39.27 | 15.31 |
| | RECAP1 | 12.89 | 12.89 | 0.965 |
| | POPDEN | 2.007 | 2.009 | 0.166 |
| | PERCHG | 10,419 | 10,389 | 840.1 |
| | LAND | 2,416 | 2,443 | 315.2 |
| | RZONED | 2,798 | 2,814 | 510.4 |
| | DRIVE | 892.8 | 888.1 | 85.68 |
| COMM | OFFICE | 2.097 | 1.362 | 0.369 |
| | HIWAYS | 13.98 | 10.78 | 4.482 |
| | ACCESS | 3,082 | 4,170 | 351.2 |
| | CZONED | 8,249 | 9,261 | 1,287 |
| | POPDEN | 0.1206 | 0.2480 | 0.0490 |
| OFFICE | COMM | 0.1149 | 0.1420 | 0.0140 |
| | HIWAYS | 2.904 | 2.155 | 1.131 |
| | JOBS | 0.1853 | 0.1660 | 0.0170 |
| | INCOME | 265.8 | 273.5 | 41.28 |
| | OZONED | 22,133 | 21,561 | 2,390 |
| MANF | HIWAYS | 16.38 | 16.05 | 3.893 |
| | PEAK | 111.3 | 112.1 | 22.89 |
| | MANJOB | 3.754 | 3.759 | 0.2830 |
| | RRMILE | 1,984 | 1,978 | 334.4 |
| | ACCESS | 3,627 | 3,585 | 919.0 |
| | GOVT | -2.201 | -2.186 | 0.7200 |
| | JOBCHG | 5.969 | 6.033 | 2.283 |
| HIWAYS | RES | 0.001452 | 0.0010 | <.0001 |
| | COMM | 0.007517 | 0.0080 | 0.0010 |
| | ACCESS | -29.75 | -29.82 | 8.215 |
| EDUC | RES | 0.03587 | 0.0360 | 0.0020 |
| | KIDS | 8.059 | 8.060 | 0.3130 |
| | STAY | 1,447 | 1,448 | 87.82 |
| | VALUE | 0.03636 | 0.0360 | 0.0050 |
| | POPDEN | 0.07166 | 0.0720 | 0.0060 |

*Shown to four significant digits, wherever possible.

## TABLE 4-7 (Continued)

## COMPARISON OF PATH REGRESSION COEFFICIENTS FROM THE FINAL PATH ANALYSIS AND THE JACKKNIFING STABILITY ANALYSIS

| Model Equation | Independent Variable | Path Regression Coefficients* | | |
|---|---|---|---|---|
| | | Final Path Analysis | Jackknifing | |
| | | | Mean | Standard Deviation |
| REC | RES | 0.007844 | 0.0080 | 0.0010 |
| | POOR | -805.0 | -805.0 | 68.13 |
| | POPDEN | 0.04382 | 0.0440 | 0.0040 |
| | POPDIF | 229.2 | 228.2 | 22.75 |
| WHOLE | WWJOB | 9.022 | 9.015 | 0.8550 |
| | ACCESS | 1,422 | 1,406 | 289.1 |
| | HIWAYS | 5.151 | 5.211 | 0.8290 |
| OTHER | REC | 1.407 | 1.394 | 0.1690 |
| | MANF | 0.1176 | 0.1170 | 0.0200 |
| | INCOME | 430.1 | 430.6 | 49.13 |

*Shown to four significant digits, wherever possible.

The percent of instability in each coefficient can be compared with the average imprecision introduced by errors inherent in the land use data on which the model is based. A previous analysis of the errors associated with estimating land use from aerial photographs concluded the average estimate to be 87% accurate (with actual values varying widely depending on the interpreter's skill, the photograph's scale and the classification system used). Introducing this error estimate (±13%) for the land use variables into the model equations produces an average imprecision of 13% for all non-endogenous variable coefficients. Comparison of this value with the 15% total variance estimated by the jackknifing indicates that aerial photo errors may be one of the principal causes of instabilities in the model coefficients.

One-way frequency distributions of the coefficient values are graphed in Appendix I for the nine model equations. To more accurately quantify the stability of the path regression coefficients and to allow comparisons, between equations, two separate indices of stability were developed.

### 1. Extreme Value Index

The first stability index that was used is a statistical test for extreme values in a sample developed by Dixon [126]. The null hypothesis assumes the value farthest from the mean does not differ significantly from the range of values in the data set, i.e., is not an extreme value. Assumptions are made that the data are independent observations and normally distributed. The coefficient values that were tested in the current application, however, are a dependent set. The violation of this assumption, however, does not invalidate our use of the test statistic as a guide for judging stability. The 40 values for each model coefficient were ordered from lowest to highest $X_1 < X_2 < \dots < X_{39} < X_{40}$, and the following ratio was computed:

$$r_1 = \frac{X_3 - X_1}{X_{k-2} - X_1}, \quad k = 40$$

The critical $r_1$ level tested for was 0.414, corresponding to the one percent significance level. If $r_1$ exceeds 0.414, then given an independent sample, there is a 99 percent probability that the deviation of the lowest value $X_1$ from the range of values in the data set is significant and that $X_1$ is an extreme value. Again as a guide, if $r_1$ exceeds 0.414 in our samples, we will refer to the corresponding $X_1$ as an extreme value significant at the .01 level. A separate test on the highest data value $X_{40}$ was also performed. A second ratio to test the highest data point was computed as follows:

$$r_{40} = \frac{X_k - X_{k-2}}{X_k - X_3}, \quad k = 40$$

The values of $r_1$ and $r_{40}$ for each model coefficient are shown in Table 4-8. Where the ratio was found to be significant at the one percent level, the case number of the data sample excluded from the path analysis corresponding to the extreme value is indicated.

When an extreme value occurs in the data set of path regression coefficient values, it indicates that the exclusion of data from a particular case number is causing a significant change in the value of the coefficient. Normally one would test for a change in value when certain data were included, however, it was impractical to compute coefficients for all possible data combinations, and thus, detection of changes was accomplished by excluding each sample, one at a time. If exclusion of one sample from a least squares analysis causes a significant change in a coefficient's values, then that data sample, when included, is exerting a greater effect on the analysis results than other samples.

The repetition of case study numbers in a given model equation in Table 4-8 (e.g. 18 appears 3 times in the OTHER equation), and the correspondence between many of the case study numbers with those in Table 4-4 indicates that skewness in the dependent variable distributions is the principal cause of model coefficient instabilities. Thus, the extreme value indices are limited to assessing the overall stability of an equation, not its individual coefficients. In this light, only the COMM, MANF, WHOLE and OTHER equations can be classified as having appreciable instability in their

## TABLE 4-8

## EXTREME VALUE INDEX OF STABILITY FOR ALL PATH REGRESSION COEFFICIENTS IN THE CAUSAL MODEL EQUATIONS

| Model Equation | Independent Variable | Extreme Value Ratio | | Case Number Excluded if Extreme Value* | |
|---|---|---|---|---|---|
| | | $r_1$ | $r_{40}$ | $X_1$ | $X_{40}$ |
| RES | HIWAYS | 0.338 | 0.190 | | |
| | RECAP1 | 0.541 | 0.064 | 34 | |
| | NONHH | 0.551 | 0.246 | 40 | |
| | POPDEN | 0.525 | 0.255 | 7 | |
| | PERCHG | 0.525 | 0.212 | 11 | |
| | LAND | 0.128 | 0.113 | | |
| | RZONED | 0.504 | 0.277 | 37 | |
| | DRIVE | 0.761 | 0.260 | 7 | |
| COMM | OFFICE | 0.662 | 0.709 | 35 | 3 |
| | HIWAYS | 0.221 | 0.339 | | |
| | ACCESS | 0.366 | 0.442 | | 20 |
| | CZONED | 0.578 | 0.432 | 11 | 29 |
| | POPDEN | 0.616 | 0.332 | 3 | |
| OFFICE | COMM | 0.184 | 0.498 | | 35 |
| | HIWAYS | 0.342 | 0.399 | | |
| | JOBS | 0.354 | 0.087 | | |
| | INCOME | 0.684 | 0.185 | 17 | |
| | OZONED | 0.800 | 0.415 | 3 | 8 |
| MANF | HIWAYS | 0.642 | 0.305 | 18 | |
| | PEAK | 0.688 | 0.560 | 35 | 9 |
| | MANJOB | 0.219 | 0.492 | | 30 |
| | RRMILE | 0.545 | 0.677 | 30 | 34 |
| | ACCESS | 0.207 | 0.632 | | 30 |
| | GOVT | 0.270 | 0.592 | | 34 |
| | JOBCHG | 0.464 | 0.658 | 30 | 18 |
| HIWAYS | RES | 0.385 | 0.196 | | |
| | COMM | 0.076 | 0.353 | | |
| | ACCESS | 0.292 | 0.429 | | |
| EDUC | RES | 0.446 | 0.354 | 11 | |
| | KIDS | 0.277 | 0.493 | | 29 |
| | STAY | 0.471 | 0.300 | 3 | |
| | VALUE | 0.796 | 0.708 | 3 | 17 |
| | POPDEN | 0.402 | 0.183 | | |

*At the 1 percent significance level.

## TABLE 4-8 (Continued)

## EXTREME VALUE INDEX OF STABILITY FOR ALL PATH REGRESSION COEFFICIENTS IN THE CAUSAL MODEL EQUATIONS

| Model Equation | Independent Variable | Extreme Value Ratio $r_1$ | $r_{40}$ | Case Number Excluded if Extreme Value* $X_1$ | $X_{40}$ |
|---|---|---|---|---|---|
| REC | RES | 0.204 | 0.446 | | 7 |
| | POOR | 0.476 | 0.395 | 3 | |
| | POPDEN | 0.305 | 0.241 | | |
| | POPDIF | 0.200 | 0.099 | | |
| WHOLE | WWJOB | 0.480 | 0.616 | 14 | 17 |
| | ACCESS | 0.888 | 0.484 | 20 | 18 |
| | HIWAYS | 0.528 | 0.716 | 14 | 18 |
| OTHER | REC | 0.853 | 0.404 | 18 | |
| | MANF | 0.860 | 0.852 | 18 | 35 |
| | INCOME | 0.659 | 0.641 | 18 | 17 |

*At the 1 percent significance level.

model estimates. These are the only equations having greater than 50 percent extreme values. Since multiple occurrences of the same case number in a given equation is not as strong an indication of instability as the occurrence of the same number of different case numbers, the COMM and WHOLE equations can be classified as having the greatest instabilities.

### 2. Coefficient of Variation Index

A secondary stability index was developed to quantify the effect of the total variation in each model coefficient on the predicted values for the dependent variable. The variation in each path regression coefficient was expressed as half of the difference between the maximum ($b_{max}$) and minimum ($b_{min}$) values obtained in the jackknifing analysis. This difference was then multiplied by the mean value of the associated independent variable ($\overline{X}_i$) to obtain the total effect on the dependent variable. This quantity was then normalized by the mean value of the associated dependent variable ($\overline{Y}$) to obtain the percentage change:

$$P_i = \left| \frac{(b_{max} - b_{min})\overline{X}_i}{2\overline{Y}} \right|$$

The computed $P_i$ values are summarized in Table 4-9.

Of the 43 values of $P_i$ listed in Table 4-9, only 5 of them (or 12%) exceed 30% and only 2 of them (or 5%) exceed 40%. The average value of $P_i$ is 0.16 (16%). Thus, no widespread instability problem exists and the large instabilities that do occur are restricted to the COMM and OTHER equations.

## TABLE 4-9

## STABILITY INDEX FOR ALL PATH REGRESSION COEFFICIENTS IN THE CAUSAL MODEL EQUATIONS BASED ON THE EFFECT OF VARIATION IN THE COEFFICIENTS ON THE DEPENDENT VARIABLE

| Model Equation | Independent Variable | Percentage Change in Dependent Variable |
|---|---|---|
| RES | HIWAYS | 0.21 |
| | RECAP1 | 0.06 |
| | NONHH | 0.02 |
| | POPDEN | 0.13 |
| | PERCHG | 0.07 |
| | LAND | 0.05 |
| | RZONED | 0.06 |
| | DRIVE | 0.06 |
| COMM | OFFICE | 0.41 |
| | HIWAYS | 0.30 |
| | ACCESS | 0.05 |
| | CZONED | 0.08 |
| | POPDEN | 0.19 |
| OFFICE | COMM | 0.16 |
| | HIWAYS | 0.28 |
| | JOBS | 0.07 |
| | INCOME | 0.28 |
| | OZONED | 0.09 |
| MANF | HIWAYS | 0.23 |
| | PEAK | 0.22 |
| | MANJOB | 0.07 |
| | RRMILE | 0.23 |
| | ACCESS | 0.12 |
| | GOVT | 0.07 |
| | JOBCHG | 0.13 |
| HIWAYS | RES | 0.10 |
| | COMM | 0.15 |
| | ACCESS | 0.06 |
| EDUC | RES | 0.08 |
| | KIDS | 0.15 |
| | STAY | 0.19 |
| | VALUE | 0.34 |
| | POPDEN | 0.05 |

TABLE 4-9 (Continued)

STABILITY INDEX FOR ALL PATH REGRESSION COEFFICIENTS IN THE CAUSAL MODEL EQUATIONS BASED ON THE EFFECT OF VARIATION IN THE COEFFICIENTS ON THE DEPENDENT VARIABLE

| Model Equation | Independent Variable | Percentage Change in Dependent Variable |
|---|---|---|
| REC | RES | 0.15 |
| | POOR | 0.17 |
| | POPDEN | 0.11 |
| | POPDIF | 0.05 |
| WHOLE | WWJOB | 0.17 |
| | ACCESS | 0.19 |
| | HIWAYS | 0.27 |
| OTHER | REC | 0.27 |
| | MANF | 0.34 |
| | INCOME | 0.49 |

## C. DETERMINE NET CAUSAL EFFECTS

### 1. Approach

The objectives of this task were to quantify direct and indirect causal effects between model variables, and to summarize the total net relationships. This causal tracing analysis was performed for every variable interaction in the final path analytic model. The total net effects obtained in this manner correspond to the path coefficients of the "reduced-form" equations [15]. An introduction to causal tracing is given in Section 8 of Appendix A.

In general, the total effect (T) from an input variable (X) to a dependent variable (Y) is determined by:

$$T = \left[ \frac{(E_1 + E_2 + \ldots\ldots)\,(1-L_1)(1-L_2)(1-L_3)\ldots\ldots}{(1-L_1)(1-L_2)(1-L_3)\ldots\ldots} \right]^*$$

where: $E_1$, $E_2$, ..... are distinct open paths from X to Y

$L_1$, $L_2$, ..... are the return effects for all distinct loops providing relevant feedback

* is a special operation in which the multiplications in the numerator and in the denominator are carried out before division, terms are deleted if they multiply the effects of touching paths, and division is carried out only after such terms have been deleted.

A more thorough description of the procedure is given by Heise [15]. An example of this technique is illustrated below for the input variable OFFICE and the dependent variable OTHER. Referring to Figure 4-1, it can be seen there are two open paths:

$E_1$ : OFFICE →COMM →HIWAYS →RES →REC →OTHER

$E_1 = 0.50 * 0.51 * 0.21 * 0.25 * 0.42 = 0.00562$

$E_2$ : OFFICE →COMM →HIWAYS →MANF →OTHER

$E_2 = 0.50 * 0.51 * 0.21 * 0.45 = 0.0241$

These open paths touch four feedback loops:

$L_1$ : OFFICE ↔ COMM

$L_1 = 0.49 * 0.50 = 0.245$

$L_2$ : COMM ↔ HIWAYS

$L_2 = 0.21 * 0.51 = 0.107$

$L_3$ : HIWAYS ↔ RES

$L_3 = 0.27 * 0.21 = 0.0567$

$$L_4 : \begin{tikzcd} \text{OFFICE} \arrow[r] & \text{COMM} \arrow[r] & \text{HIWAYS} \arrow[ll, bend right] \end{tikzcd}$$

$L_4 = 0.50 * 0.51 * 0.18 = 0.0459$

Thus, the total net effect is:

$$T = \left[ \frac{(E_1 + E_2)(1-L_1)(1-L_2)(1-L_3)(1-L_4)}{(1-L_1)(1-L_2)(1-L_3)(1-L_4)} \right]^*$$

Carrying out the multiplication in the denominator, all of the cross terms are deleted except for $L_1L_3$ since these loops do not touch. Carrying out the multiplication in the numerator, all of the E and L cross terms are deleted since every feedback loop touches $E_1$ and $E_2$. The resultant expression is:

$$T = \frac{E_1 + E_2}{(1-L_1-L_2-L_3-L_4+L_1L_3)}$$

$$= 0.05$$

### 2. Results and Conclusions

The direct, indirect and total causal effects between the 9 endogenous land use variables and all input variables are listed in Table 4-10. through 4-18. Conclusions, for each model equation, are given below by summarizing the 5 most important causal factors, in descending order, with a prefix sign indicating the direction of causal action. Note that a discussion of the reasons for including each variable that appears in the model is given in Section II.H.

#### (1) RES Equation

The most important causal factors for residential development are measures of (+) population characteristics, (+) accessibility, (+) regional growth, (+) reserve collection capacity of the wastewater

## TABLE 4-10

## SUMMARY OF DIRECT AND INDIRECT CAUSAL EFFECTS ON THE VARIABLE RES

| Causal Factor |  | Path Coefficients |  |  |
|---|---|---|---|---|
| Type | Variable | Direct | Indirect | Total |
| Land Use Variables | COMM | -- | 0.19 | 0.19 |
|  | OFFICE | -- | 0.10 | 0.10 |
|  | MANF | -- | -- | -- |
|  | HIWAYS | 0.21 | 0.02 | 0.23 |
|  | EDUC | -- | -- | -- |
|  | REC | -- | -- | -- |
|  | WHOLE | -- | -- | -- |
|  | OTHER | -- | -- | -- |
| Population Characteristics | POPDEN | 0.48 | 0.05 | 0.53 |
|  | NONHH | -0.19 | -0.01 | -0.20 |
|  | KIDS | -- | -- | -- |
| Employment Characteristics | JOBS | -- | 0.02 | 0.02 |
|  | MANJOB | -- | -- | -- |
|  | WWJOB | -- | -- | -- |
| Accessibility | ACCESS | -- | 0.0005 | 0.0005 |
|  | DRIVE | 0.41 | 0.03 | 0.44 |
|  | RRMILE | -- | -- | -- |
| Regional Influences | POPDIF | -- | -- | -- |
|  | PERCHG | 0.40 | 0.03 | 0.43 |
|  | JOBCHG | -- | -- | -- |
| Income | INCOME | -- | 0.01 | 0.01 |
|  | POOR | -- | -- | -- |
| Major Project | RECAP1 | 0.32 | 0.02 | 0.34 |
|  | PEAK | -- | -- | -- |
| Zoning | RZONED | 0.13 | 0.01 | 0.14 |
|  | CZONED | -- | 0.03 | 0.03 |
|  | OZONED | -- | 0.04 | 0.04 |
| Mobility | STAY | -- | -- | -- |
| Housing Characteristics | VALUE | -- | -- | -- |
| Land Costs, Taxes | LAND | 0.22 | 0.02 | 0.24 |
|  | GOVT | -- | -- | -- |

## TABLE 4-11

## SUMMARY OF DIRECT AND INDIRECT CAUSAL EFFECTS ON THE VARIABLE COMM

| Causal Factor |  | Path Coefficients |  |  |
|---|---|---|---|---|
| Type | Variable | Direct | Indirect | Total |
| Land Use Variables | RES | -- | 0.06 | 0.06 |
|  | OFFICE | 0.50 | 0.34 | 0.84 |
|  | MANF | -- | -- | -- |
|  | HIWAYS | 0.21 | 0.14 | 0.35 |
|  | EDUC | -- | -- | -- |
|  | REC | -- | -- | -- |
|  | WHOLE | -- | -- | -- |
|  | OTHER | -- | -- | -- |
| Population Characteristics | POPDEN | 0.08 | 0.12 | 0.20 |
|  | NONHH | -- | -0.02 | -0.02 |
|  | KIDS | -- | -- | -- |
| Employment Characteristics | JOBS | -- | 0.17 | 0.17 |
|  | MANJOB | -- | -- | -- |
|  | WWJOB | -- | -- | -- |
| Accessibility | ACCESS | 0.18 | 0.08 | 0.26 |
|  | DRIVE | -- | 0.06 | 0.06 |
|  | RRMILE | -- | -- | -- |
| Regional Influences | POPDIF | -- | -- | -- |
|  | PERCHG | -- | 0.06 | 0.06 |
|  | JOBCHG | -- | -- | -- |
| Income | INCOME | -- | 0.10 | 0.10 |
|  | POOR | -- | -- | -- |
| Major Project | RECAP1 | -- | 0.05 | 0.05 |
|  | PEAK | -- | -- | -- |
| Zoning | RZONED | -- | 0.02 | 0.02 |
|  | CZONED | 0.16 | 0.11 | 0.27 |
|  | OZONED | -- | 0.30 | 0.30 |
| Mobility | STAY | -- | -- | -- |
| Housing Characteristics | VALUE | -- | -- | -- |
| Land Costs, Taxes | LAND | -- | 0.03 | 0.03 |
|  | GOVT | -- | -- | -- |

## TABLE 4-12

## SUMMARY OF DIRECT AND INDIRECT CAUSAL EFFECTS ON THE VARIABLE OFFICE

| Causal Factor | | Path Coefficients | | |
|---|---|---|---|---|
| Type | Variable | Direct | Indirect | Total |
| | RES | -- | 0.14 | 0.14 |
| | COMM | 0.49 | 0.41 | 0.90 |
| Land Use Variables | MANF | -- | -- | -- |
| | HIWAYS | 0.18 | 0.30 | 0.48 |
| | EDUC | -- | -- | -- |
| | REC | -- | -- | -- |
| | WHOLE | -- | -- | -- |
| | OTHER | -- | -- | -- |
| Population Characteristics | POPDEN | -- | 0.14 | 0.14 |
| | NONHH | -- | -0.03 | -0.03 |
| | KIDS | -- | -- | -- |
| Employment Characteristics | JOBS | 0.20 | 0.10 | 0.30 |
| | MANJOB | -- | -- | -- |
| | WWJOB | -- | -- | -- |
| Accessibility | ACCESS | -- | 0.12 | 0.12 |
| | DRIVE | -- | 0.06 | 0.06 |
| | RRMILE | -- | -- | -- |
| Regional Influences | POPDIF | -- | -- | -- |
| | PERCHG | -- | 0.06 | 0.06 |
| | JOBCHG | -- | -- | -- |
| Income | INCOME | 0.12 | 0.06 | 0.18 |
| | POOR | -- | -- | -- |
| Major Project | RECAP1 | -- | 0.05 | 0.05 |
| | PEAK | -- | -- | -- |
| Zoning | RZONED | -- | 0.02 | 0.02 |
| | CZONED | -- | 0.16 | 0.16 |
| | OZONED | 0.35 | 0.17 | 0.52 |
| Mobility | STAY | -- | -- | -- |
| Housing Characteristics | VALUE | -- | -- | -- |
| Land Costs, Taxes | LAND | -- | 0.03 | 0.03 |
| | GOVT | -- | -- | -- |

## TABLE 4-13
## SUMMARY OF DIRECT AND INDIRECT CAUSAL EFFECTS ON THE VARIABLE MANF

| Causal Factor Type | Variable | Path Coefficients Direct | Indirect | Total |
|---|---|---|---|---|
| Land Use Variables | RES | -- | 0.08 | 0.08 |
| | COMM | -- | 0.14 | 0.14 |
| | OFFICE | -- | 0.10 | 0.10 |
| | HIWAYS | 0.21 | -- | 0.21 |
| | EDUC | -- | -- | -- |
| | REC | -- | -- | -- |
| | WHOLE | -- | -- | -- |
| | OTHER | -- | -- | -- |
| Population Characteristics | POPDEN | -- | 0.05 | 0.05 |
| | NONHH | -- | -0.01 | -0.01 |
| | KIDS | -- | -- | -- |
| Employment Characteristics | JOBS | -- | 0.02 | 0.02 |
| | MANJOB | 0.31 | -- | 0.31 |
| | WWJOB | -- | -- | -- |
| Accessibility | ACCESS | 0.18 | 0.0005 | 0.18 |
| | DRIVE | -- | 0.03 | 0.03 |
| | RRMILE | 0.32 | -- | 0.32 |
| Regional Influences | POPDIF | -- | -- | -- |
| | PERCHG | -- | 0.02 | 0.02 |
| | JOBCHG | 0.09 | -- | 0.09 |
| Income | INCOME | -- | 0.01 | 0.01 |
| | POOR | -- | -- | -- |
| Major Project | RECAP1 | -- | 0.02 | 0.02 |
| | PEAK | 0.22 | -- | 0.22 |
| Zoning | RZONED | -- | 0.01 | 0.01 |
| | CZONED | -- | 0.03 | 0.03 |
| | OZONED | -- | 0.03 | 0.03 |
| Mobility | STAY | -- | -- | -- |
| Housing Characteristics | VALUE | -- | -- | -- |
| Land Costs, Taxes | LAND | -- | 0.01 | 0.01 |
| | GOVT | -0.08 | -- | -0.08 |

## TABLE 4-14

## SUMMARY OF DIRECT AND INDIRECT CAUSAL EFFECTS ON THE VARIABLE HIWAYS

| Causal Factor | | Path Coefficients | | |
|---|---|---|---|---|
| Type | Variable | Direct | Indirect | Total |
| Land Use Variables | RES | 0.27 | 0.09 | 0.36 |
| | COMM | 0.51 | 0.18 | 0.69 |
| | OFFICE | -- | 0.46 | 0.46 |
| | MANF | -- | -- | -- |
| | EDUC | -- | -- | -- |
| | REC | -- | -- | -- |
| | WHOLE | -- | -- | -- |
| | OTHER | -- | -- | -- |
| Population Characteristics | POPDEN | -- | 0.25 | 0.25 |
| | NONHH | -- | -0.07 | -0.07 |
| | KIDS | -- | -- | -- |
| Employment Characteristics | JOBS | -- | 0.09 | 0.09 |
| | MANJOB | -- | -- | -- |
| | WWJOB | -- | -- | -- |
| Accessibility | ACCESS | -0.12 | 0.12 | 0.002 |
| | DRIVE | -- | 0.15 | 0.15 |
| | RRMILE | -- | -- | -- |
| Regional Influences | POPDIF | -- | -- | -- |
| | PERCHG | -- | 0.14 | 0.14 |
| | JOBCHG | -- | -- | -- |
| Income | INCOME | -- | 0.06 | 0.06 |
| | POOR | -- | -- | -- |
| Major Project | RECAP1 | -- | 0.12 | 0.12 |
| | PEAK | -- | -- | -- |
| Zoning | RZONED | -- | 0.05 | 0.05 |
| | CZONED | -- | 0.15 | 0.15 |
| | OZONED | -- | 0.16 | 0.16 |
| Mobility | STAY | -- | -- | -- |
| Housing Characteristics | VALUE | -- | -- | -- |
| Land Costs, Taxes | LAND | -- | 0.08 | 0.08 |
| | GOVT | -- | -- | -- |

## TABLE 4-15
## SUMMARY OF DIRECT AND INDIRECT CAUSAL EFFECTS ON THE VARIABLE EDUC

| Causal Factor | | Path Coefficients | | |
|---|---|---|---|---|
| Type | Variable | Direct | Indirect | Total |
| Land Use Variables | RES | 0.48 | -- | 0.48 |
| | COMM | -- | 0.07 | 0.07 |
| | OFFICE | -- | 0.05 | 0.05 |
| | MANF | -- | -- | -- |
| | HIWAYS | -- | 0.11 | 0.11 |
| | REC | -- | -- | -- |
| | WHOLE | -- | -- | -- |
| | OTHER | -- | -- | -- |
| Population Characteristics | POPDEN | 0.23 | 0.26 | 0.49 |
| | NONHH | -- | -0.09 | -0.09 |
| | KIDS | 0.43 | -- | 0.43 |
| Employment Characteristics | JOBS | -- | 0.01 | 0.01 |
| | MANJOB | -- | -- | -- |
| | WWJOB | -- | -- | -- |
| Accessibility | ACCESS | -- | 0.0002 | 0.0002 |
| | DRIVE | -- | 0.20 | 0.20 |
| | RRMILE | -- | -- | -- |
| Regional Influences | POPDIF | -- | -- | -- |
| | PERCHG | -- | 0.19 | 0.19 |
| | JOBCHG | -- | -- | -- |
| Income | INCOME | -- | 0.006 | 0.006 |
| | POOR | -- | -- | -- |
| Major Project | RECAP1 | -- | 0.15 | 0.15 |
| | PEAK | -- | -- | -- |
| Zoning | RZONED | -- | 0.06 | 0.06 |
| | CZONED | -- | 0.01 | 0.01 |
| | OZONED | -- | 0.02 | 0.02 |
| Mobility | STAY | 0.34 | -- | 0.34 |
| Housing Characteristics | VALUE | 0.26 | -- | 0.26 |
| Land Costs, Taxes | LAND | -- | 0.11 | 0.11 |
| | GOVT | -- | -- | -- |

## TABLE 4-16

## SUMMARY OF DIRECT AND INDIRECT CAUSAL EFFECTS ON THE VARIABLE REC

| Causal Factor | | Path Coefficients | | |
|---|---|---|---|---|
| Type | Variable | Direct | Indirect | Total |
| Land Use Variables | RES | 0.25 | -- | 0.25 |
| | COMM | -- | 0.04 | 0.04 |
| | OFFICE | -- | 0.02 | 0.02 |
| | MANF | -- | -- | -- |
| | HIWAYS | -- | 0.06 | 0.06 |
| | EDUC | -- | -- | -- |
| | WHOLE | -- | -- | -- |
| | OTHER | -- | -- | -- |
| Population Characteristics | POPDEN | 0.33 | 0.13 | 0.46 |
| | NONHH | -- | -0.05 | -0.05 |
| | KIDS | -- | -- | -- |
| Employment Characteristics | JOBS | -- | 0.005 | 0.005 |
| | MANJOB | -- | -- | -- |
| | WWJOB | -- | -- | -- |
| Accessibility | ACCESS | -- | 0.0001 | 0.0001 |
| | DRIVE | -- | 0.11 | 0.11 |
| | RRMILE | -- | -- | -- |
| Regional Influences | POPDIF | 0.23 | -- | 0.23 |
| | PERCHG | -- | 0.11 | 0.11 |
| | JOBCHG | -- | -- | -- |
| Income | INCOME | -- | 0.002 | 0.002 |
| | POOR | -0.27 | -- | -0.27 |
| Major Project | RECAP1 | -- | 0.09 | 0.09 |
| | PEAK | -- | -- | -- |
| Zoning | RZONED | -- | 0.03 | 0.03 |
| | CZONED | -- | 0.008 | 0.008 |
| | OZONED | -- | 0.014 | 0.014 |
| Mobility | STAY | -- | -- | -- |
| Housing Characteristics | VALUE | -- | -- | -- |
| Land Costs, Taxes | LAND | -- | 0.06 | 0.06 |
| | GOVT | -- | -- | -- |

## TABLE 4-17

## SUMMARY OF DIRECT AND INDIRECT CAUSAL EFFECTS ON THE VARIABLE WHOLE

| Causal Factor | | Path Coefficients | | |
|---|---|---|---|---|
| Type | Variable | Direct | Indirect | Total |
| Land Use Variables | RES | -- | 0.10 | 0.10 |
| | COMM | -- | 0.19 | 0.19 |
| | OFFICE | -- | 0.12 | 0.12 |
| | MANF | -- | -- | -- |
| | HIWAYS | 0.27 | -- | 0.27 |
| | EDUC | -- | -- | -- |
| | REC | -- | -- | -- |
| | OTHER | -- | -- | -- |
| Population Characteristics | POPDEN | -- | 0.07 | 0.07 |
| | NONHH | -- | -0.02 | -0.02 |
| | KIDS | -- | -- | -- |
| Employment Characteristics | JOBS | -- | 0.02 | 0.02 |
| | MANJOB | -- | -- | -- |
| | WWJOB | 0.41 | -- | 0.41 |
| Accessibility | ACCESS | 0.29 | 0.0006 | 0.29 |
| | DRIVE | -- | 0.04 | 0.04 |
| | RRMILE | -- | -- | -- |
| Regional Influences | POPDIF | -- | -- | -- |
| | PERCHG | -- | 0.04 | 0.04 |
| | JOBCHG | -- | -- | -- |
| Income | INCOME | -- | 0.01 | 0.01 |
| | POOR | -- | -- | -- |
| Major Project | RECAP1 | -- | 0.03 | 0.03 |
| | PEAK | -- | -- | -- |
| Zoning | RZONED | -- | 0.01 | 0.01 |
| | CZONED | -- | 0.04 | 0.04 |
| | OZONED | -- | 0.04 | 0.04 |
| Mobility | STAY | -- | -- | -- |
| Housing Characteristics | VALUE | -- | -- | -- |
| Land Costs, Taxes | LAND | -- | 0.02 | 0.02 |
| | GOVT | -- | -- | -- |

## TABLE 4-18

## SUMMARY OF DIRECT AND INDIRECT CAUSAL EFFECTS ON THE VARIABLE OTHER

| Causal Factor | | Path Coefficients | | |
|---|---|---|---|---|
| Type | Variable | Direct | Indirect | Total |
| Land Use Variables | RES | -- | 0.14 | 0.14 |
| | COMM | -- | 0.07 | 0.07 |
| | OFFICE | -- | 0.05 | 0.05 |
| | MANF | 0.45 | -- | 0.45 |
| | HIWAYS | -- | 0.12 | 0.12 |
| | EDUC | -- | -- | -- |
| | REC | 0.42 | -- | 0.42 |
| | WHOLE | -- | -- | -- |
| Population Characteristics | POPDEN | -- | 0.22 | 0.22 |
| | NONHH | -- | -0.03 | -0.03 |
| | KIDS | -- | -- | -- |
| Employment Characteristics | JOBS | -- | 0.01 | 0.01 |
| | MANJOB | -- | 0.14 | 0.14 |
| | WWJOB | -- | -- | -- |
| Accessibility | ACCESS | -- | 0.08 | 0.08 |
| | DRIVE | -- | 0.06 | 0.06 |
| | RRMILE | -- | 0.14 | 0.14 |
| Regional Influences | POPDIF | -- | 0.10 | 0.10 |
| | PERCHG | -- | 0.05 | 0.05 |
| | JOBCHG | -- | 0.04 | 0.04 |
| Income | INCOME | 0.15 | 0.006 | 0.16 |
| | POOR | -- | -0.11 | -0.11 |
| Major Project | RECAP1 | -- | 0.04 | 0.04 |
| | PEAK | -- | 0.10 | 0.10 |
| Zoning | RZONED | -- | 0.02 | 0.02 |
| | CZONED | -- | 0.01 | 0.01 |
| | OZONED | -- | 0.02 | 0.02 |
| Mobility | STAY | -- | -- | -- |
| Housing Characteristics | VALUE | -- | -- | -- |
| Land Costs, Taxes | LAND | -- | 0.03 | 0.03 |
| | GOVT | -- | -0.04 | -0.04 |

major project and (+) land costs. It is interesting to note that variables representing the major project treatment plant capacity were not significant in the causal model, i.e., it is the collection network not the treatment facilities which can be said to induce residential development. This is consistent with the discussion of wastewater major project impacts in Sections I.B.3 and II.E. Other factors, which were hypothesized in the initial model, but which turned out not to be important causal factors, are manufacturing development, outdoor recreational development, on-site disposal restrictions, income, employment characteristics and vacant developable land. Vacant developable land is a prerequisite for induced development. The reason it did not appear to be causally important is probably because all case study areas were chosen a priori to include significant amounts of vacant land.

### (2) COMM Equation

The most important causal factors for commercial development are measures of (+) office development, (+) non-expressway highway lane miles, (+) office zoning, (+) commercial zoning, and (+) expressway interchange density. Major project timing was hypothesized as a causal factor, but the analysis rejected this variable. However, collection network reserve capacity is a small indirect causal factor (through residential development). All other hypothesized causal factors were verified in some measure, large or small.

### (3) OFFICE Equation

The most important causal factors for office-professional development are measures of (+) commercial development, (+) office zoning, (+) non-expressway highway lane miles, (+) employment density, and (+) family income. Again, an hypothesized relationship with major project timing was rejected by the analysis. However, collection network reserve capacity is a small indirect causal factor (through residential development). Other causal factors which turned out not to be important were manufacturing development and development constraints.

### (4) MANF Equation

The most important causal factors for manufacturing development are measures of (+) railroad mileage density, (+) manufacturing employment density, (+) wastewater major project peak flow, (+) non-expressway highway lane miles, and (+) expressway interchange density. Although a measure of wastewater system capacity was found to be significant, it is interesting to note that measures of sewer service (area of land served) and major project timing were rejected by the analysis. Other causal factors which turned out not to be important were on-site disposal restrictions and vacant developable land.

### (5) HIWAYS Equation

The most important causal factors for non-expressway highway development are (+) commercial development, (+) office development, (+) residential development, (+) population density, and (+) office zoning. Although the hypothesized relationship with major project timing was rejected by the analysis, collection network reserve capacity is an indirect causal factor (through residential development). Other causal factors which turned out to be unimportant were manufacturing development and distance to trip generators.

### (6) EDUC Equation

The most important causal factors for educational development are (+) population density, (+) residential development, (+) school-age children per household, (+) neighborhood stability, and (+) residential real estate values. Again, although the hypothesized relationship with major project timing was rejected, collection network reserve capacity is an indirect causal factor (through residential development). All other hypothesized causal factors were verified in some measure, large or small.

### (7) REC Equation

The most important causal factors for active, outdoor recreational devleopment are (+) population density, (-) percent of families below poverty level, (+) residential development, (+) growth in regional population, and (+) growth in county population. All other hypothesized causal factors were verified, with the exception of educational development.

### (8) WHOLE Equation

The most important causal factors for wholesale-warehouse development are (+) wholesale and warehouse employment density, (+) expressway interchange density, (+) non-expressway highway lane miles, (+) commercial development, and (+) office development. Although the hypothesized relationship with major project timing was rejected, collection network reserve capacity is a small indirect causal factor (through residential development). All other hypothesized causal factors were verified in some measure, with the exception of manufacturing development and vacant developable land.

### (9) OTHER Equation

The most important causal factors for other types of development are (+) manufacturing development, (+) recreational development, (+) population density, (+) family income, and (+) residential development. Although major project timing was rejected as a causal factor, wastewater system peak flow is an indirect causal factor (through manufacturing development. All other hypothesized causal factors were verified, with the exception of distance to population center.

## D. VMT MODEL VALIDATION

The objective of this task was to validate the GEMLUP VMT model by: (1) testing its predictive ability with transportation data from 11 to the 40 case studies, and (2) quantifying the standard error associated with its use in these cases. The VMT model is a methodology to predict the vehicle miles travelled by motor vehicles in trips associated with land use within either the region or the area of analysis. The basic technique for predicting vehicle miles of travel (VMT) is simply to form the product of: (1) predictions of land use, (2) vehicle trip generation rates, and (3) vehicle trip lengths, and sum over all land use categories. Thus, it was our intent not only to validate the final VMT projection, but the process of obtaining trip generation rates and trip lengths as well (either by equation or default values). Our approach to this task involved first developing criteria for determining whether model performance in the validation exercise is "reasonable", i.e., are the errors encountered of the same order as those encountered in other simplified VMT predictive techniques and are they excessive in comparison to typical errors introduced at other points in the GEMLUP predictive model. These acceptability criteria are discussed below.

### 1. Develop Test Criteria

Two convenient measures of a model's accuracy and precision in any validation are the standard error of estimate and correlation coefficient values. Thus, in developing criteria, the literature was first searched for validations of similar transportation models in which these statistics had been quantified, for the three key GEMLUP parameters: (1) VMT, (2) trip generation rates, and (3) trip lengths. The most extensive compilation of data on existing transportation procedures and models is given in Appendix B of a recent publication by Comsis Corporation [127]. Although the effectiveness and limitations of 40 different procedures are discussed, these evaluations are unfortunately all subjective and non-numeric. Four other publications were found that did contain significant data.

The Maricopa Association of Governments has produced a summary report of data from several hundred land use trip generation studies done nationwide[120]. The data are arranged in linear relationships between land use (dwelling units or 1000 square feet gross floor area) and trip generation rates. Regression analysis is then used to test the accuracy of the concept of linear predictive equations for trip rates. Applicable data from this publication are shown in Table 4-19. Bates [128] describes the result of a validation analysis of trip generation equations, developed from data on 4 southeastern U.S. cities and tested with data from 8 others. Standard errors of estimate for VMT ranged from -21.1% to +24.5% (average ±22.8%), while for trip generation rates the errors were -53.6% to +70.0% (average ±61.8%). A study by Jeffries and Carter [129] of trip generation in small urban areas used data from 6 states to test simple trip rate equations for home-based trips. Correlation results ranged from 0.74 to 0.95 with an overall coefficient value of 0.86. Finally, Ashford and Holloway [130] did a comprehensive validation of trip production equations developed for Pittsburgh, PA. Their results are summarized in Table 4-20.

Based on these data, it was decided to use the following as testing criteria:

- For VMT, the maximum allowable error is ±23%.
- For trip generation rates, the maximum allowable errors and minimum correlations to be used are summarized in Table 4-21. Where exact data on a given land category were not available, overall average statistics for trip rate prediction of 42% and R = 0.81 were used.

The existing literature on validation provides data on only two of the three key GEMLUP parameters, viz. VMT and trip rates. For trip lengths a different approach was required to estimate reasonable errors of estimation. It was assumed that errors should be some percentage of the total variation typically seen in a parameter. If errors approach 100% of the total range, it is probable that a value above the mean could be estimated as below the mean, or vice-versa. Thus, a value of 50% of the total variation in trip length data was used as the maximum allowable error. Only one publication was found that gave substantial data on the variation of average trip lengths. The FHWA[131] has compiled extensive data on trip

TABLE 4-19

COMPILATION OF NATIONWIDE DATA ON THE RELATIONSHIP BETWEEN TRIP GENERATION RATES AND QUANTITIES OF LAND USE

| Trip Generation Land Use Type | No. of Studies | Average Standard Deviation of Trip Rates (%) | Average Linear Correlation (R) |
|---|---|---|---|
| Residential | | | |
| Single Family | 231 | 32% | 0.88 |
| Multi-Family | 65 | 9% | 0.95 |
| All | 296 | 21% | 0.92 |
| Commercial | | | |
| Small retail | 58 | 52% | 0.61 |
| Shopping centers | | | |
| Small | 54 | 50% | 0.56 |
| Medium | 24 | 16% | 0.89 |
| Large | 18 | 19% | 0.90 |
| All | 144 | 34% | 0.74 |
| Offices | 15 | 36% | 0.78 |
| Manufacturing | | | |
| Small | 39 | 77% | 0.26 |
| Large | 15 | 25% | 0.87 |
| All | 54 | 51% | 0.57 |
| Warehousing | 16 | 103% | 0.83 |
| Average Statistics | | 49% | 0.74 |

Source: [120].

## TABLE 4-20

## VALIDATION RESULTS FOR TRIP PRODUCTION EQUATIONS DEVELOPED FOR PITTSBURGH, PA

| Trip Type | Average Error | Correlation Coefficient |
|---|---|---|
| Home-Based | | |
| Work | +15.4% | 0.97 |
| Shopping | -18.7% | 0.89 |
| School | -28.0% | 0.70 |
| Miscellaneous | + 9.0% | -- |
| Non-work | - 6.5% | -- |
| Other | - 1.7% | -- |
| Nonhome-Based | -19.7% | 0.76 |
| Average Statistics | ±14.1% | 0.83 |

Source: [130].

## TABLE 4-21

## SUGGESTED VALIDATION CRITERIA FOR TRIP GENERATION RATES

| Land Use Category | Maximum Error (±%) | Minimum Correlation (R) |
|---|---|---|
| Residential | | |
| Single family detached | 32% | 0.88 |
| Single family attached | 32% | 0.88 |
| Multi-family low rise | 9% | 0.95 |
| Multi-family high rise | 9% | 0.95 |
| Mobile homes | 32% | 0.88 |
| Hotels/Motels | 42% | 0.81 |
| Commercial | | |
| <50K | 51% | 0.59 |
| 50-100K | 16% | 0.89 |
| >100K | 19% | 0.90 |
| Office | 36% | 0.78 |
| Manufacturing | 51% | 0.57 |
| Wholesale/Warehousing | 103% | 0.83 |
| Cultural | 42% | 0.81 |
| Churches | 42% | 0.81 |
| Hospitals | 42% | 0.81 |
| Educational | 28% | 0.70 |
| Recreation | 42% | 0.81 |

length distributions by various parameters, such as trip purpose, size of urban area and family income. The range in these data are summarized in Table 4-22. Average statistics extracted from these published data suggest the following as testing criteria:

- For trip lengths, the maximum allowable errors to be used are 44% for work trips and 40% for other trips.

The suggested maximum error criteria are not excessive in comparison to typical errors introduced elsewhere in the GEMLUP-I predictive model. The average error of the land use predictive equations is ±87% [19]. The EPA emission factors used in the translation from land use to emissions are averages for source classes and when applied to individual sources can have errors ranging from one to several hundred percent [132]. Finally, diffusion models, which are ultimately used to predict air quality levels, are precise to only a factor of 2 or 3 in general [133].

### 2. Model Validation

Three separate validations were performed to test the predictive ability of the:

- Predictive equations for trip lengths
- Default values for trip generation rates
- VMT model as a whole using as input default values for trip rates and the default predictive equation for trip lengths.

The structure of default values and predictive equations in the existing VMT model dictated the level of detail in the validation. Thus, trip lengths were averaged for the area of analysis by trip purpose only (work or other). Trip generation rates were averages for the area of analysis by land use type and purpose, and VMT were totals for the area of analysis. A fourth validation, involving VMT estimates based on local trip rate and trip length data, could not be performed due to lack of sufficient, detailed data from the transportation case studies. The validation of each of these three elements is discussed in the sections below.

## TABLE 4-22

## TRIP LENGTH DATA DISTRIBUTIONS COMPILED BY THE FEDERAL HIGHWAY ADMINISTRATION

| Distribution Parameters | Trip Length Total (miles) | Range as Percentage of Mean |
|---|---|---|
| Size of Urban Area | | |
| Earning a Living | 4.1 | 40% |
| Family Business | 3.2 | 57% |
| Civic, Educational & Religious | 3.5 | 74% |
| Social & Recreational | 8.6 | 66% |
| Family Income | | |
| Earning a Living | 11.3 | 111% |
| Family Business | 7.9 | 141% |
| Civic, Eduational & Religious | 2.6 | 55% |
| Social & Recreational (excluding vacations) | 14.8 | 113% |
| Average Statistics | | |
| Work Trips | 6.6 | 87% |
| Other Trips | 7.4 | 77% |

Source: [131].

### a. Trip Lengths

Actual data on trip lengths were available from 8 case studies for "work trips" and from 7 case studies for "other trips". The GEMLUP VMT model was used to predict estimates of "work" and "other" trip lengths, given data on network speeds and SMSA population. These data are summarized in Table 4-23 and scattergrams of actual versus predicted values are given in Figures 4-2 and 4-3. Correlation coefficients are 0.49 and 0.03, respectively. for work and other trips. Both predictive equations generally underestimate trip length, with an average bias of -32%. The standard errors of estimate are 50% and 56%, respectively, for work and other trips. These exceed the acceptable criteria limits of 44% and 40%, respectively. Thus, it was necessary to revise the predictive equations for trip length.

The technique used to revise the trip length equations was to adjust the multiplicative constant in each case. The adjustment factor was found by fitting the data to a regression line with a forced zero intercept and computing the slope of the line. These regression lines are drawn in Figures 4-2 and 4-3 and have slopes of 1.49 and 1.46, respectively. Thus, the new predictive equations become:

$$L_1 = 0.00447 * P^{0.20} * S_1^{1.49}$$

$$L_2 = 0.00219 \left[ P^{0.18} * S_2^{1.40} + P^{0.26} * S_2^{1.25} \right]$$

The bias of these equations, applied to the case study data, is zero. The imprecision can be characterized by the revised standard errors of estimate.

The revised model was tested with the case study data and standard errors of estimate of 34% and 39% were obtained, respectively. As these do not exceed the acceptable criteria limits, the revised trip length equations were judged suitable for use in the GEMLUP-II model.

The residual errors from the second validation are graphed versus the actual trip length data in Figures 4-4 and 4-5. Examination of these graphs show no evidence of changing variance in the error term, i.e., no heteroscedasticity.

# TABLE 4-23

# ACTUAL AND PREDICTED TRIP LENGTH DATA IN MILES

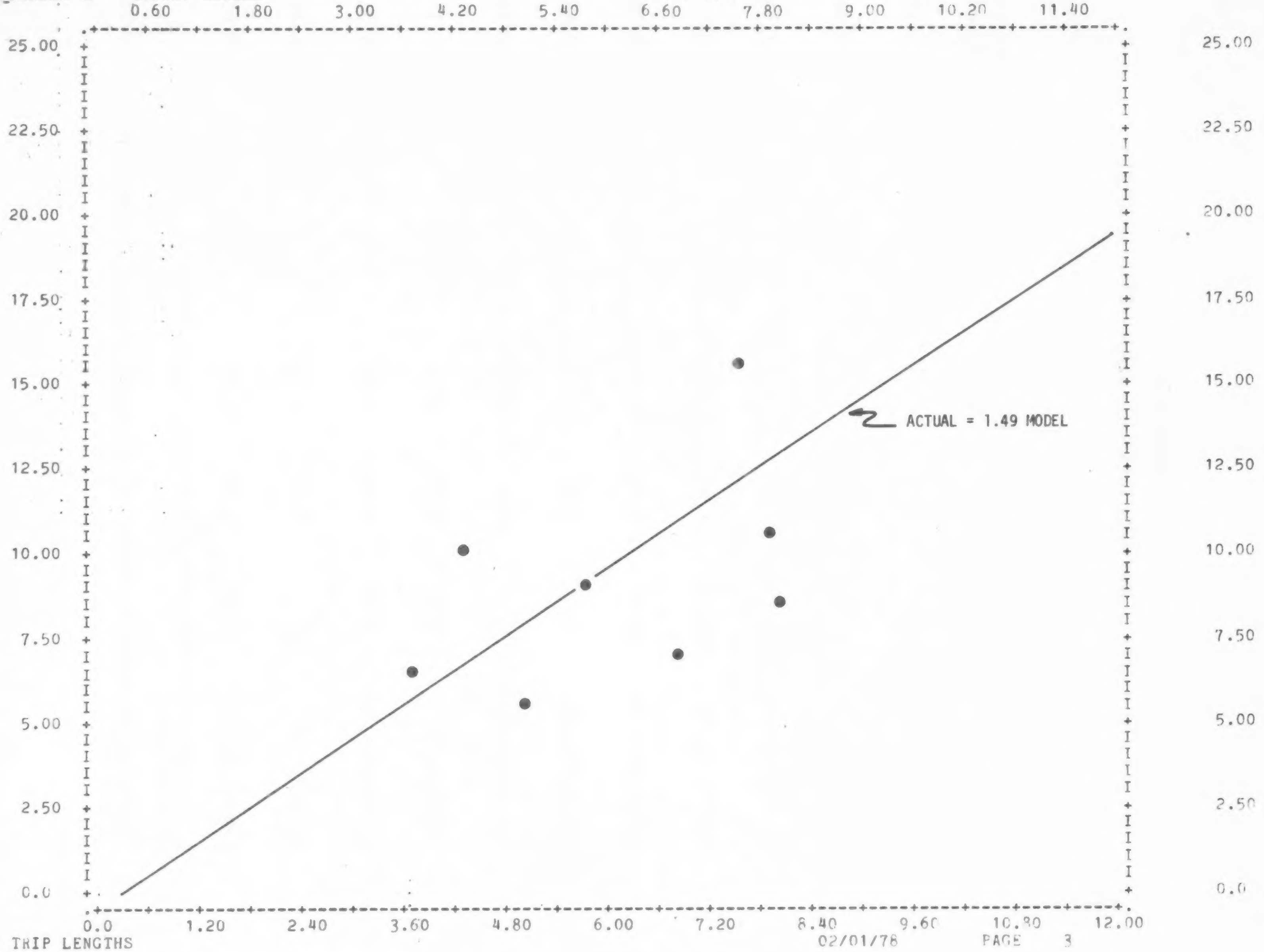
| Transportation Case Study | WORK TRIPS Actual | WORK TRIPS VMT Model | WORK TRIPS Revised Model | OTHER TRIPS Actual | OTHER TRIPS VMT Model | OTHER TRIPS Revised Model |
|---|---|---|---|---|---|---|
| Lexington, MA | ND | 8.79 | 13.10 | ND | 4.43 | 6.47 |
| Bennington, VT | 6.40* | 3.73 | 5.56 | 7.79* | 3.86 | 5.64 |
| Richmond, VA | 5.70 | 5.05 | 7.52 | 5.70 | 5.10 | 7.45 |
| Charlotte, NC | 10.73* | 7.95 | 11.85 | ND | 5.90 | 8.61 |
| Clearwater, FL | 15.44 | 7.57 | 11.28 | 11.05 | 5.95 | 8.69 |
| Roseville, MN | 6.91 | 6.87 | 10.24 | 5.09 | 5.09 | 7.43 |
| St. Louis, MO | ND | 11.28 | 16.81 | ND | 8.85 | 12.92 |
| Boulder, CO | 9.76 | 4.30 | 6.41 | 11.12 | 4.91 | 7.17 |
| S. Phoenix, AZ | 9.16 | 5.72 | 8.52 | 7.82 | 5.89 | 8.60 |
| Vallejo, CA | ND | 5.72 | 8.52 | ND | 4.31 | 6.29 |
| Auburn, WA | 8.43 | 8.10 | 12.07 | 5.79 | 6.44 | 9.40 |

ND = No Data

*Estimated by weighted interzonal distances from origin/destination studies circa 1970.

FIGURE 4-2
SCATTERGRAM OF ACTUAL VERSUS PREDICTED WORK TRIP LENGTHS

FILE NUMBERS (CREATION DATE = 02/01/78)
SCATTERGRAM OF (DOWN) ACTUAL (ACROSS) MODEL



WORK TRIP LENGTHS 02/01/78 PAGE 3

FIGURE 4-3
SCATTERGRAM OF ACTUAL VERSUS PREDICTED OTHER TRIP LENGTHS

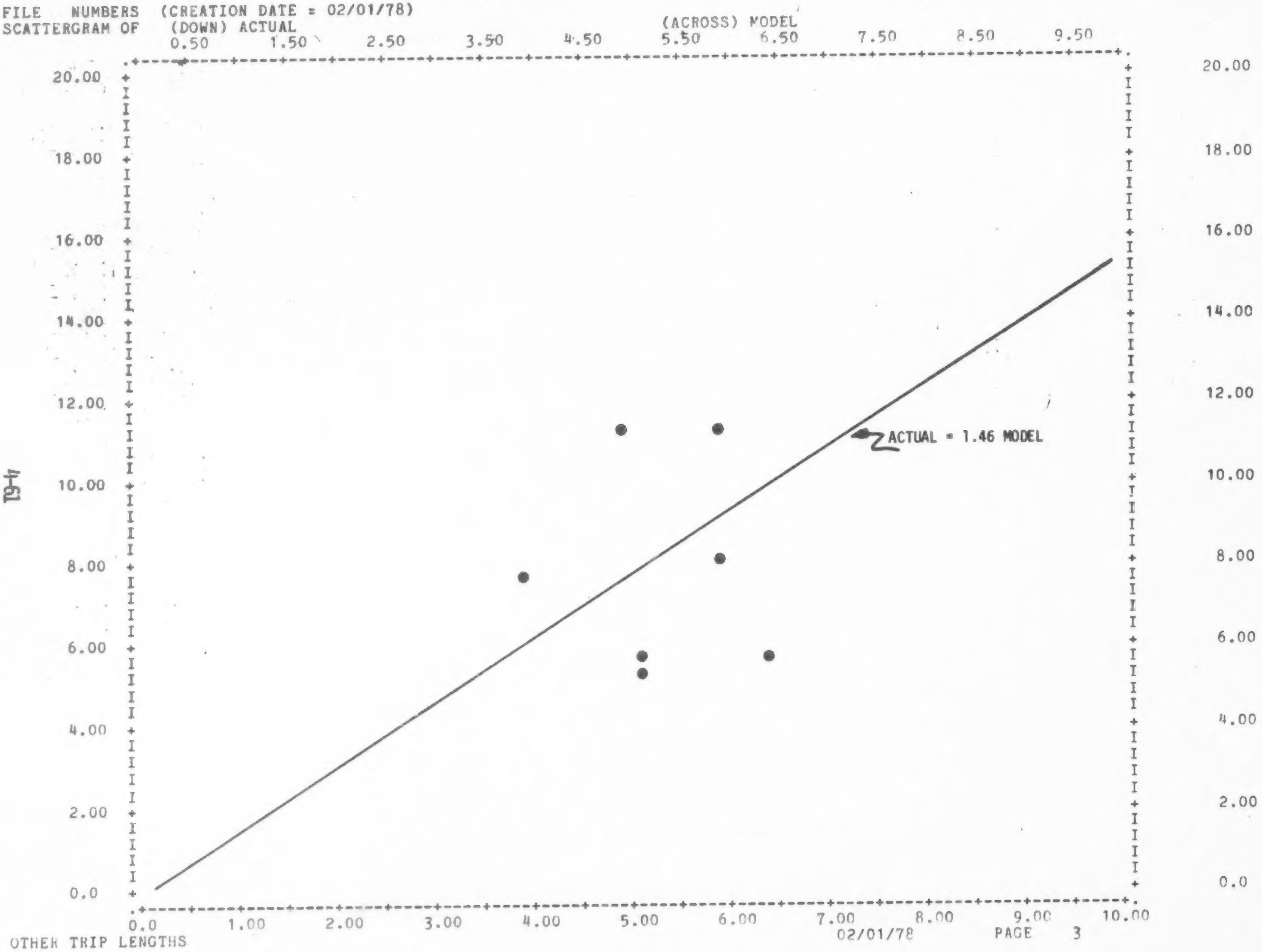


FIGURE 4-4
SCATTERGRAM OF ACTUAL WORK TRIP LENGTHS VERSUS MODEL RESIDUALS

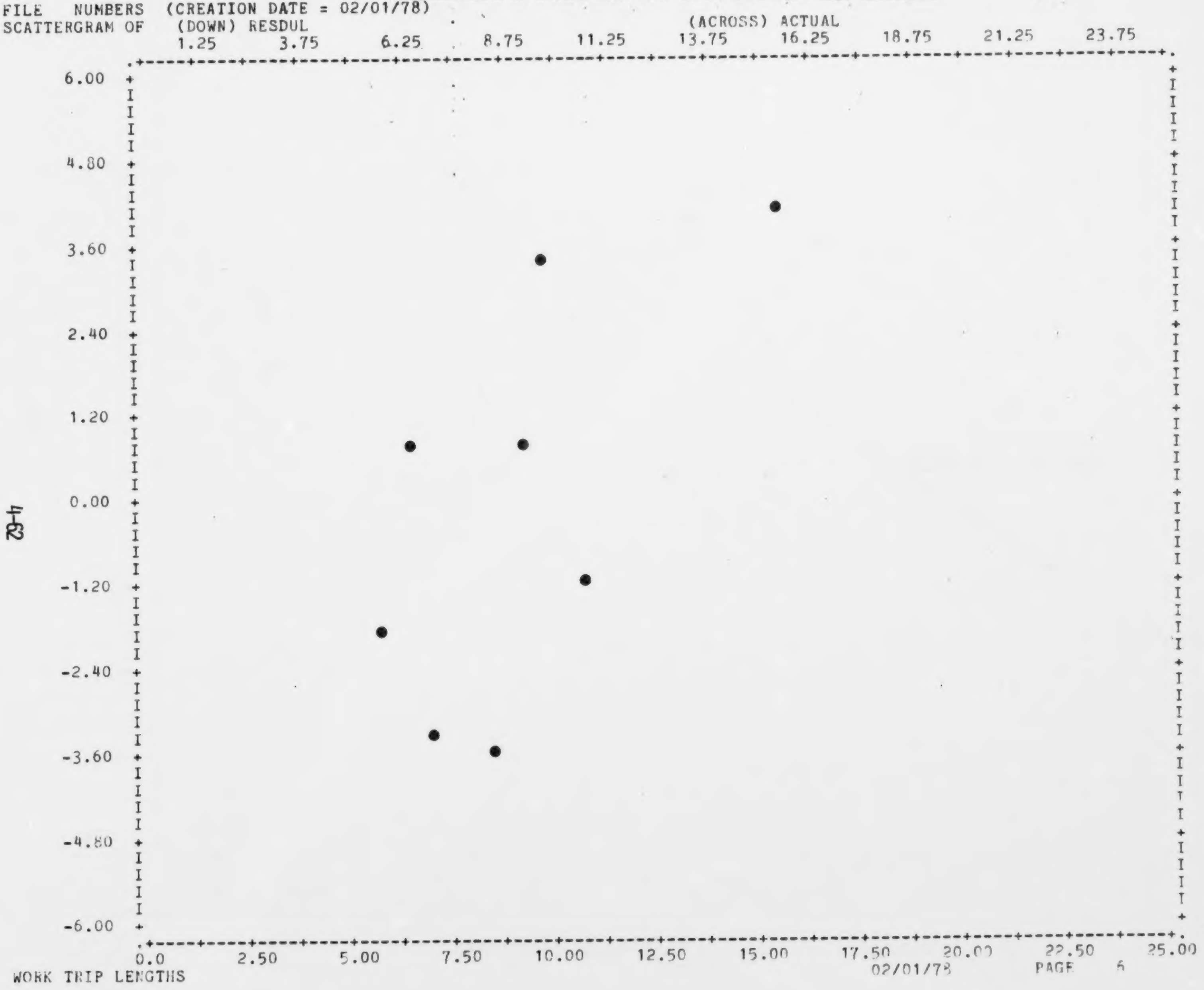
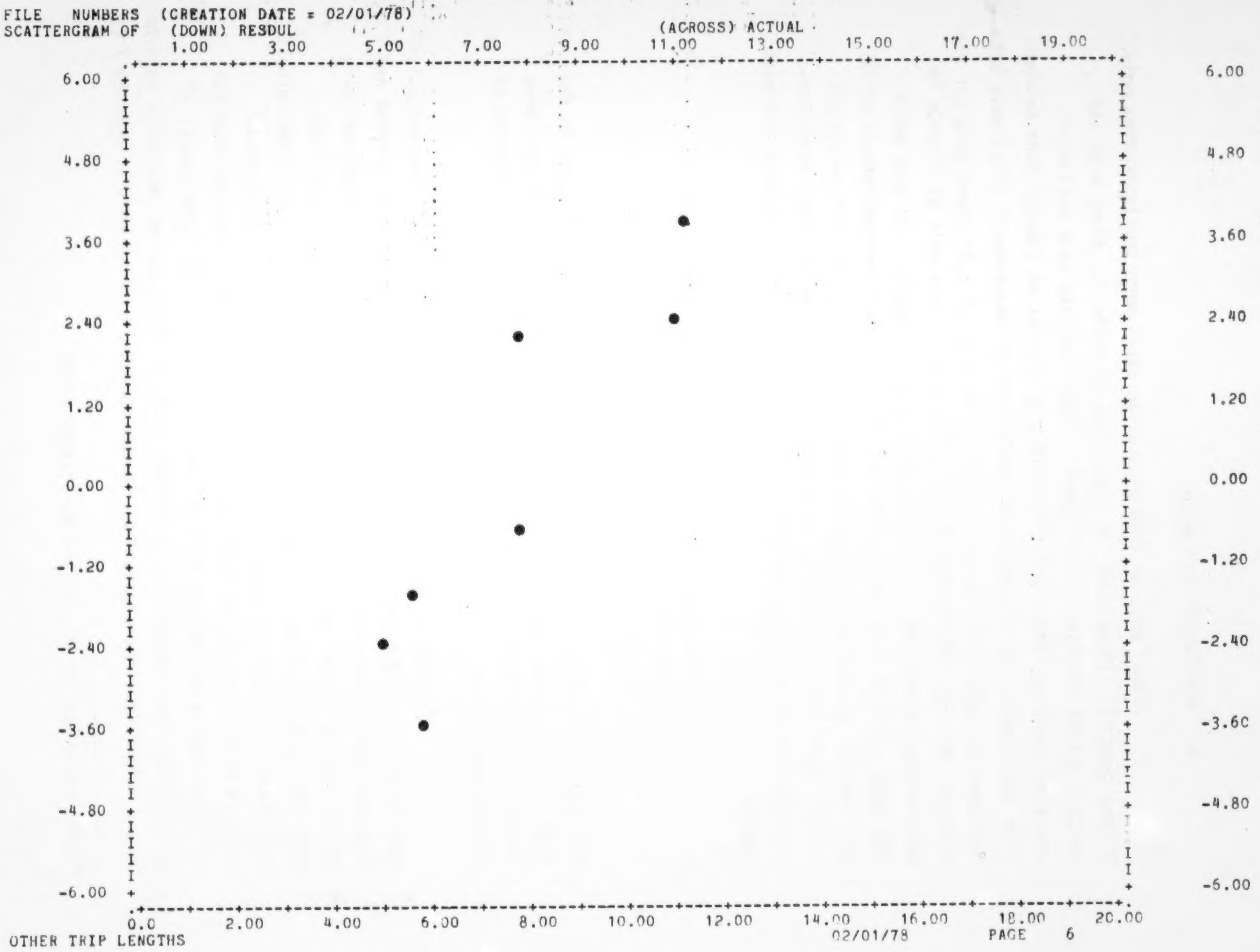


FIGURE 4-5
SCATTERGRAM OF ACTUAL OTHER TRIP LENGTHS VERSUS MODEL RESIDUALS



### b. Trip Generation Rates

Actual data on trip generation rates were available from only 2 case studies. These data are summarized in Table 17, along with the default values used in the VMT model. Since few data were available, averages computed from several hundred trip generation studies done nationwide and reported by the Maricopa Association of Governments [120] have been included in Table 4-24, as well. Due to the lack of sufficient data for statistical testing, the maximum error criteria were used only as a guide in determining whether VMT model trip rates are reasonable. On this basis, the rates for the first two commercial categories are overestimated, while the rate for recreation is underestimated. All other default values are valid. Revisions made to the trip rate values in the VMT model were performed so as to bring the default values into line with nationwide averages, namely:

- Commercial, <50K GLA = 67 trips/$10^3 ft^2$
- Commercial, 50-100K GLA = 64 trips/$10^3 ft^2$
- Recreation = 42 trips/acre

Note that recreation trip rates are the most variable of all those in the model. Thus, it may not be possible for a single average (no matter what its value) to accurately represent the trip generation characteristics of recreational areas.

Although actual trip rate data were scarce, most transportation case studies provided data on total trips, by some breakdown of purpose and land use type. Therefore, similar total trip estimates were prepared using the revised VMT model applied to collected land use data.* Comparisons were then made, in order to further validate (in an indirect way) the default trip rates. For a given case study, total daily trips for the area of analysis were estimated by summing trip production and attraction data for all zones approximating the area of analysis. Due to the large number of calculations involved and the different bases on which actual data were obtained (e.g., origin/destination studies, cordon surveys, Gravity model), and the fact

---

*Model estimates do not include duplicated trips.

## TABLE 4-24

## COMPARISON OF ACTUAL AND PREDICTED TRIP GENERATION RATES BY LAND USE

| Land Use Type/ Trip Purpose | Trips Per Measure | Auburn, WA | St. Louis, MO | Nationwide Average[120] | VMT Model |
|---|---|---|---|---|---|
| Residential/Work | Dwelling Unit | | | | |
| SF Detached | | 1.7** | -- | -- | 1.8 |
| SF Attached | | -- | -- | -- | 1.5 |
| MF Low Rise | | -- | -- | -- | 1.2 |
| MF High Rise | | -- | -- | -- | 0.8 |
| Mobile Homes | | -- | -- | -- | 1.8 |
| Residential/Other | Dwelling Unit | | | | |
| SF Detached | | 4.4** | 8.0* | 9.5* | 9.0 |
| SF Attached | | -- | 7.0* | 7.0* | 7.0 |
| MF Low Rise | | -- | 7.0* | 7.0* | 6.0 |
| MF High Rise | | -- | 7.0* | -- | 4.0 |
| Mobile Homes | | -- | 5.5* | 6.1* | 5.0 |
| Commercial, <50K GLA | $10^3 ft^2$ | | 25-80 | 67 | 130 |
| Commercial, 50-100K GLA | $10^3 ft^2$ | | 25-80 | 64 | 80 |
| Commercial, >100K GLA | $10^3 ft^2$ | | 30-65 | 34-46 | 40 |
| Office | $10^3 ft^2$ | | 12-20 | 10-25 | 16 |
| Manufacturing | $10^3 ft^2$ | | 5.0 | 4.2-9.3 | 5.0 |
| Wholesale/Warehousing | $10^3 ft^2$ | | 9.0 | 5.5 | 4.0 |
| Culture | $10^3 ft^2$ | | -- | -- | 2.0 |
| Churches | $10^3 ft^2$ | | -- | -- | 2.0 |
| Hotel/Motel | $10^3 ft^2$ | | -- | -- | 10 |
| Hospitals | $10^3 ft^2$ | | -- | -- | 16 |
| Education | $10^3 ft^2$ | | -- | -- | 4.0 |
| Recreation | Acre | | -- | 42 | 10 |

* Total for work and other trip purposes
**Average for all residential types

that the "actual" data are in reality model estimates themselves, it is expected that a large amount of imprecision exists in these data. Thus, applying maximum error criteria or minimum correlations were not appropriate However, the comparison still provided a method for estimating model bias. Due to the mis-match between the land use and trip type categories of available data from transportation agencies and those used in the VMT model, only 3 combined categories provided sufficient data for comparison. Table 4-25 summarizes the test data by case study. The average bias of the model predictions are:

-6% for Residential/Work trips
+10% for Residential/Other trips
+22% for Other Trips

These values are well within any reasonable limits for this type of model and indicate, if anything, that the model is conservative in general, i.e., it slightly overestimates total vehicle trips.

c. VMT

Direct data on total VMT were available for only 2 case studies. However, indirect estimates of VMT, based on zonal origin/destination data, were available for an additional 6 case studies, providing a total data base of 8 samples for the validation exercise. Table 4-26 summarizes these actual VMT data, along with the estimates produced by the revised VMT model. All VMT estimates used a value for $L_r$ based on a circle with an area equivalent to that of the actual area of analysis. Thus,

$$L_r = ((\text{AREA}/640)/\pi)^{1/2}$$

A scattergram of actual versus predicted VMT is given in Figure 4-6. The correlation coefficient is 0.93, indicating excellent predictive ability for the revised VMT model. The regression line of best fit has slope of 1.05 ± 0.19 and intercept of 35,792 ± 65,043. Since the standard errors of estimate for the slope and intercept exceed the differences between

## TABLE 4-25

## COMPARISON OF ACTUAL AND PREDICTED* DAILY TRIPS IN THE AREA OF ANALYSIS BY LAND USE AND PURPOSE

| Transportation Case Study | Residential |  | Other Land Use |
|---|---|---|---|
|  | Work Trip | Other Trip | All Trips |
| Lexington, MA | 7,132<br>(7,490) | 17,317<br>(22,380 | 36,202<br>(33,984) |
| Bennington,VT | ND<br>(4,590) | ND<br>(13,320) | ND<br>(89,803) |
| Richmond, VA | 40,698<br>(29,560) | 87,008<br>(88,350) | 88,946<br>(98,489) |
| Charlotte, NC | 3,250<br>(7,500) | ND<br>(22,440) | ND<br>(41,033) |
| Clearwater, FL | 24,455<br>(49,940) | 73,500<br>(148,980) | 115,849<br>(202,675) |
| Roseville, MN | 36,466<br>(21,890) | 100,731<br>(65,580) | 36,971<br>(213,784) |
| St. Louis, MD | ND<br>(95,950) | ND<br>(287,340) | ND<br>(438,149) |
| Boulder, CO | 28,729<br>(38,430) | 144,296<br>(114,720) | 184,081<br>(139,428) |
| S. Phoenix, AZ | 24,057<br>(29,700) | 43,965<br>(88,980) | 80,728<br>(23,568) |
| Vallejo, CA | 35,828<br>(8,400) | 88,420<br>(97,680) | 63,864<br>(145,802) |
| Auburn, WA | 19,467<br>(14,650) | 51,906<br>(43,860) | 125,748<br>(36,342) |

ND = No Data
*Predicted values are in parantheses.

## TABLE 4-26

## COMPARISON OF ACTUAL AND PREDICTED TOTAL VMT FOR THE AREA OF ANALYSIS

| Transportation Case Study | Total VMT | |
|---|---|---|
| | Actual | Predicted |
| Lexington, MA | 97,810* | 102,975 |
| Bennington, VT | ND | -- |
| Richmond, VA | 608,676* | 607,870 |
| Charlotte, NC | ND** | -- |
| Clearwater, FL | 1,136,800 | 1,123,422 |
| Roseville, MN | 519,746 | 898,990 |
| St. Louis, MO | ND*** | -- |
| Boulder, CO | 1,126,616* | 1,267,922 |
| S. Phoenix, AZ | 554,035* | 512,588 |
| Vallejo, CA | 484,758* | 649,090 |
| Auburn, WA | 273,559 | 175,076 |

ND = No Data

* VMT estimated from zonal origin/destination statistics and area of analysis size.

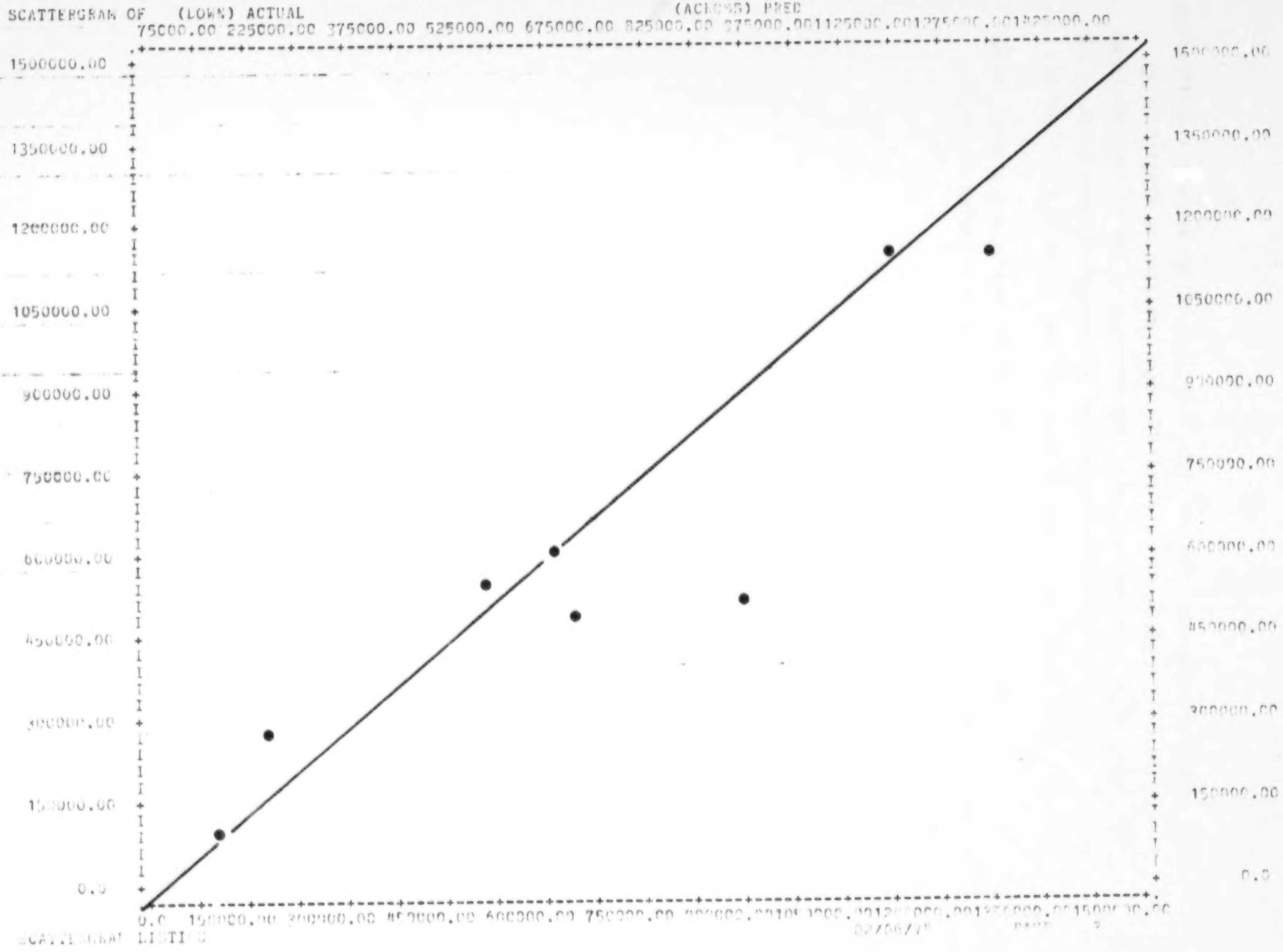
** VMT data for entire city, not just area of analysis.

***Land use data incomplete to translate trip rates to trip totals.

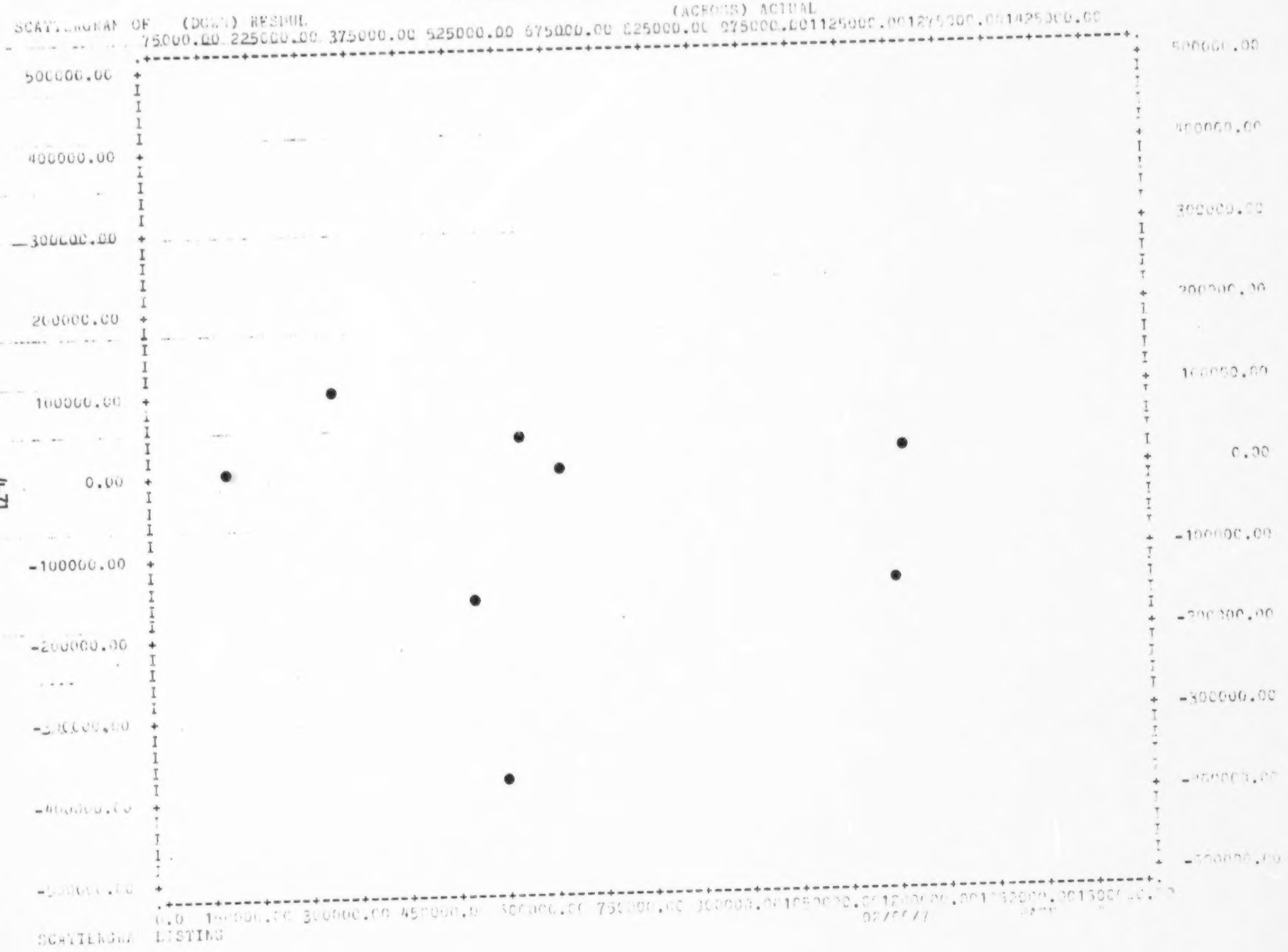
the parameter values and the set (1.0, 0.0), the values are not significantly from (1.0, 0.0). Hence, the revised VMT model contains no statistically significant bias. The average error (imprecision) of the revised VMT model is 18%. As this value does not exceed the acceptable criteria limit of 23%, the revised VMT model was judged suitable for use in the GEMLUP-II model.

The residual errors from the VMT validation are graphed versus actual VMT data in Figure 4-7. Examination shows no evidence of heteroscedasticity.

FIGURE 4-6
SCATTERGRAM OF ACTUAL VERSUS PREDICTED VMT



## FIGURE 4-7
## SCATTERGRAM OF ACTUAL VMT VERSUS MODEL RESIDUALS



## V. SUMMARY OF RESULTS AND CONCLUSIONS

A theoretical model of the Growth Effects of Major Land Use Projects has been developed. This model represents the total land use in the drainage basin of a wastewater collection and treatment system (the major project), ten years after its construction. The model represents the process of induced land use growth in the following 9 land use categories.

Residential
Commercial
Office - Professional
Manufacturing
Highways (Non-expressway)
Education
Recreation
Wholesale/Warehouse
Other

The assumption of a single basic causal structure for induced development, and the use of cross-sectional data from 40 diverse case study major projects throughout the United States, allowed a static approach to the testing of the theoretical model, using path analysis.

Path analysis is a set of statistical techniques useful in testing theories and studying the logical consequences of various hypotheses involving causal relations. It is not capable of deducing or generating causal relations, only testing them. The causal analysis of induced land use development in the current study involved the use of two basic statistical techniques: two-stage least squares and stepwise ordinary least squares (multiple regression). The first technique was required to produce consistent estimates of the path coefficients in a system of simultaneous equations involving feedback loops (or reciprocal causation) in the models. The second technique was used to solve the remaining recursive portions of the models. The dependent variables in these regression analyses represented the total land use in the previously noted 9 categories. Both linear and non-linear forms were tested and the linear form was found to produce the best fit. Specific statistical criteria were developed to identify model paths that were insignificant or redundant, and these criteria were used in an iterative process to trim unneeded and undesirable paths from the models. The trimming process eliminated almost half of the paths in the models as originally specified. The statistical problems of multicollinearity, suppressor variables and identification were eliminated through the approach used to trim the initial model.

The final models of land use development show that strong statistical relationships exist between the variables representing the 9 categories of total land use and the other model variables representing induced and non-induced land use growth processes. The results indicate that the final model explains the majority of the variance in the case study data with $R^2$ values ranging from 0.27 to 0.82 and averaging 0.54. The residuals of the final regressions do not exhibit any trends or patterns, indicating the remaining unexplained variance ($1-R^2$) is not due to poor specification of the model, but rather due to wide variance in the case study data (i.e., the problem of trying to develop one generalized model for a broad range of situations). An analysis of the stability of the model coefficients determined that, in general, the coefficient values have low variance (±15%) and exhibit no extreme instabilities. An analysis of the net causal effects in the land use model indicates that reserve capacity in the collection system of a wastewater major project is a significant causal factor for induced land use growth, principally in the residential, manufacturing, education and highways categories. It is interesting to note that treatment plant capacity was not found to be an important causal factor.

The GEMLUP-I VMT model [19] was validated using transportation data from 11 of the 40 case study major projects. Based on the validation results, it was concluded that revisions to both the default predictive equations for trip length and the default values for trip generation rates were necessary. The revised VMT model was validated and found to have an average error (imprecision) of 23%, with no statistically significant bias.

## VI. REFERENCES

1. Wright, S., "The Method of Path Coefficients", *Annals of Mathematical Statistics*, 5:161-215, 1934.
2. Wright, S., "The Interpretation of Multivariate Systems", *Statistics and Mathematics in Biology*, Edited by O. Kempthorne, T. Bancraft, J. Cowen, and J. Lush, Iowa State College Press, Ames, IA, 1954.
3. Wright, S., "Path Coefficients and Path Regressions Alternative or Complementary Concepts", *Biometrics*, 16:189, 1960.
4. Wright, S., "The Treatment of Reciprocal Interaction, with or without Lag, in Path Analysis", *Biometrics*, 16:423, 1960.
5. Van de Geer, J.P., *Introduction to Multivariate Analysis for the Social Sciencies*, W.H. Freeman, San Francisco, 1971.
6. Kerlinger, F.N., and Pedhazer, E.J., *Multiple Regression in Behavioral Research*, Holt, Rinehart and Winston, New York, NY, 1973.
7. Heise, D.R., "Problems in Path Analysis and Causal Interference", *Sociological Methodology*, edited by E.F. Borgatta, Josey-Bass, San Francisco, CA, 1969 (pp. 38-72).
8. Goldberger, A.S., *Econometric Theory*, John Wiley & Sons, New York, NY, 1964.
9. Johnston, J., *Econometric Methods*, McGraw-Hill, New York, NY, 1963.
10. Tukey, J.W., "Causation, Regression and Path Analysis", *Statistics and Mathematics in Biology*, Edited by O. Kempthorne, T.A. Bancroft, J.W. Gowen and J.L. Lush, Iowa State College Press, Ames, IA, 1954.
11. Turner, M.E., and Stevens, C.D., "The Regression Analysis of Causal Paths", *Biometrics*, 15:236-258, 1959.
12. Snedecor, G.W., and Cochran, W.G., *Statistical Methods* (6th Edition), Iowa State College Press, Ames, IA, 1967.
13. Land, K.C., "Principles of Path Analysis", *Sociological Methodology*, Edited by E.F. Borgatta, Jossey-Bass, San Francisco, 1969.
14. Boudon, R., "A Method of Linear Causal Analysis: Dependence Analysis" *American Sociological Review*, 30:365-373, 1965.
15. Heise, D., *Causal Analysis*, John Wiley & Sons, New York, NY, 1976.
16. Wonnacott, R.J. and Wonnacott, T.H., *Econometrics*, John Wiley & Sons, Inc., New York, NY, 1970.
17. Benesh, F., Guldberg, P., and D'Agostino, R., *Growth Effects of Major Land Use Projects: Volume I - Specification and Causal Analysis of Model*, EPA Publication No. EPA-450/3-76-012a, Research Triangle Park, NC, May 1976.
18. Benesh, F., *Growth Effects of Major Land Use Projects: Volume II - Compilation of Land Use Based Emission Factors*, EPA Publication No. EPA-450/3-76-012b, Research Triangle Park, NC, September 1976.

19. Benesh, F., Guldberg, P., and D'Agostino, R., *Growth Effects of Major Land Use Projects: Volume III - Summary*, EPA Publication No. EPA-450/3-76-012C, Research Triangle Park, NC, September 1976.

20. McCurdy, T., "Request for Proposal: Growth Effects of Major Land Use Projects," RFP #DU-75-C181, Research Triangle Park, NC, December 23, 1974.

21. McCurdy, T., "Request for Proposal: Growth Effects of Major Land Use Projects (Wastewater Facilities), "EPA RFP #DU-77-C007, Research Triangle Park, NC, December 3, 1976.

22. Meehan, E.J., *The Theory and Method of Political Analysis*, The Dorsey Press, Homewood, IL, 1965.

23. Abraham Kaplan, A., *The Conduct of Inquiry,* Chandler Publications, San Francisco, CA, 1964.

24. DeGroot, A.D., *Foundations of Inference and Research in the Behavorial Sciences,* Mouton & Co., The Hague, Netherlands, 1969.

25. McCurdy T., Benesh, F., Guldberg, P., and D'Agostino, R., "Application of Path Analysis to Delineate the Secondary Growth Effects of Major Land Use Projects," in Wayne R. Ott (ed.), *Environmental Modeling and Simulation,* U.S. Environmental Protection Agency, Washington, DC, 1976.

26. Bunge, M., *Causality*, Harvard University Press, Cambridge, MA, 1959.

27. Blalock, H.M., *Causal Interferences in Nonexperimental Research*, W.W. Norton Press, New York, NY, 1972.

28. Bertalanffy, L.V., *General Systems Theory,* George Braziller, New York, NY, 1968.

29. Van de Geer, J.P., *Introduction to Multivariate Analysis for the Social Sciences,* W.H. Freeman and Company, San Francisco, 1971.

30. *Webster's Seventh New Collegiate Dictionary,* G. & C. Merriam Company, Springfield, MA, 1970.

31. Mausel, P.W., Leivo, C.E., and Lewellen, M.T., "Regional Land Use Classification from Computer-Processed Satellite Data", *Journal of the American Institute of Planners,* *42*(2):153-164, April 1964.

32. Fitzpatrick, A., *Cost, Accuracy and Consistency Comparisons of Land Use Maps Made from High-Altitude Aircraft Photography and ERTS Imagery, Final Report, Central Atlantic Ecological Test Site (CARETS) Project,* U.S. Geological Survey and National Aeronautics and Space Administration, Reston, VA, September 1975.

33. Branch, M.C., *City Planning and Aerial Information*, Harvard University Press, Cambridge, MA, 1971.

34. Lindgren, D.T., *Dwelling Unit Estimation from Color Infrared Photography,* Dartmouth College, Department of Geography Paper No. 1, Hanover, NH, May 1970.

35. Collins, W.G., El-Beik, A.H., "The Acquisition of Urban Land Use Information from Aerial Photographs of the City of Leeds (Great Britain)", *Photogrammetria*, *27*:71-92, 1971.

36. Simpson, R.B., Line-Scan vs Optical Sensors for Discrimination of Built-Up Areas, Dartmouth College, Department of Geography Paper No. 2, Hanover, NH, May 1970.

37. Personal communication, Ivan Harden, Geography Applications Program, USGS, U.S. Department of the Interior, Reston, VA, May 31, 1977.

38. Personal communication, Katherine A. Fitzpatrick, Geography Applications Program, USGS, U.S. Department of the Interior, Reston, VA, June 6, 1977.

39. Avery, T.E., Interpretation of Aerial Photographs (2nd Edition), Burgess Publishing, Co., Minneapolis, MN, 1968.

40. U.S. Department of Agriculture, Aerial-Photo Interpretation in Classifying and Mapping Soils, Soil Conservation Service, U.S. Department of Agriculture, Agriculture Handbook 294, Washington, DC, October 1966.

41. Personal communication with Katherine A. Fitzpatrick, Geography Applications Program, USGS, U.S. Department of the Interior, Reston, VA, June 7, 1977.

42. Office of Air Quality Planning and Standards, Guidelines for Air Quality Maintenance Planning and Analysis; Volume 2: Plan Preparation, EPA Publication No. EPA-450/4-74-002, Research Triangle Park, NC, 1974.

43. Office of Air Quality Planning and Standards, Guidelines for Air Quality Maintenance Planning and Analysis; Volume 3: Control Strategies, EPA Publication No. EPA-450/4-74-003, Research Triangle Park, NC, 1974.

44. Office of Air Quality Planning and Standards, Guidelines for Air Quality Maintenance Planning and Analysis; Volume 4: Land Use and Transportation Considerations, EPA Publication No. EPA-450/4-74-004,

45. Office of Air Quality Planning and Standards, Guidelines for Air Quality Maintenance Planning and Analysis; Volume 6: Overview of Air Quality Maintenance Area Analysis, EPA Publication No. EPA-450/4-74-007, Research Triangle Park, NC, 1974.

46. Office of Air Quality Planning and Standards, Guidelines for Air Quality Maintenance Planning and Analysis; Volume 9: Evaluating Indirect Sources, EPA Publication No. EPA-450/4-75-001, Research Triangle Park, NC, 1975.

47. Office of Air Quality Planning and Standards, Guidelines for Air Quality Maintenance Planning and Analysis; Volume 12: Applying Atmospheric Simulation Models to Air Quality Maintenance Areas, EPA Publication No. EPA-450/4-74-013, Research Triangle Park, NC, 1974.

48. National Environmental Policy Act of 1969, 42 U.S.C., Section 4321 et seq.

49. Council on Environmental Quality, "Guidelines for Preparation of Environmental Impact Statements," Federal Register, 38(147), Part II, August 1, 1973.

50. Personal communication, Mr. Bruce Bane, Assistant City Engineer, Los Altos, CA, December 3, 1976.

51. Personal communication, Mr. Arthur Vondrick, Water and Sewers Director, Phoenix, AZ, December 8, 1976.

52. Grants Administration Division, Project Register - Wastewater Treatment Construction Grants, U.S. Environmental Protection Agency, Washington, DC, June 30, 1973.

53. Promise, J. and Leiserson, M., Water Resources Management for Metropolitan Washington: Analysis of the Joint Interactions of Water and Sewer Service, Public Policy and Land Development Patterns in an Expanding Metropolitan Area (including Appendices), Metropolitan Washington Council of Governments, Washington, DC, 1973.

54. Grava, S., Urban Planning Aspects of Water Pollution Control, Columbia University Press, New York, NY, 1969.

55. Personal communication, Mr. Robert Richmond, Tri-State Regional Planning Commission, New York, NY, December 8, 1976.

56. U.S. Census Bureau, "Finances of Special Districts," 1972 Census of Governments, 4(2), Washington, DC, 1972.

57. Milgram, G., The City Expands - A Study of the Conversion of Land from Rural to Urban Use in Philadelphia, 1945-62, University of Pennsylvania, Philadelphia, PA, March, 1967.

58. Rabe, F.T. and Hudson, J.F., "Highway and Sewer Impacts on Urban Development," J. Urban Planning and Development - ASCE, 101:217, November 1975.

59. Tabors, R., Shapiro, M., and Rogers, P., Land Use and the Pipe: Planning for Sewerage, Heath, Lexington, MA, 1976.

60. Stansbury, J., "Suburban Growth - A Case Study," Population Bulletin, 28:5, February 1972.

61. Bascom , S.E., et al., Secondary Impacts of Transporation and Wastewater Investments: Review and Bibliography, EPA Publication No. EPA-600-5-75-002, Washington, DC, 1975.

62. Council on Environmental Quality, The Fifth Annual Report of the Council on Environmental Quality, Washington, DC, 1974.

63. Urban Systems Research and Engineering, Inc., Interceptor Sewers and Suburban Sprawl: The Impact and Construction Grants on Residential Land Use, Volume I: Analysis, and Volume II: Case Studies, prepared for the Council on Environmental Quality, Washington, DC, September 1974.

64. Urban Systems Research & Engineering, Inc., The Growth Shapers - The Land Use Impacts of Infrastructure Investments, prepared for the Council on Environmental Quality, Washington, DC, May 1976.

65. Hudson, J.F., Demand for Municipal Services: Measuring the Effect of Service Quality, Ph.D. thesis, MIT Department of Civil Engineering, Cambridge, MA, June 1975.

66. Bascom , S.E., Cooper, K.G., Howell, M.P., Makrides, A.C., and Rabe, F.T., Secondary Impacts of Transportation and Wastewater Investments: Research Results, EPA Publication No. EPA-600/5-75-013, Washington, DC, 1975.

67. Real Estate Research Corporation, The Costs of Sprawl - Environmental and Economic Costs of Alternative Residential Development Patterns at the Urban Fringe, prepared for the Council on Environmental Quality, Washington, DC, April 1974.

68. Phillips, M.B., "Developments in Water Quality and Land Use Planning: Problems in the Application of the Federal Water Pollution Control Act Amendments of 1972", Urban Law Annual, 10:43, 1975.

69. Personal communication, Dr. Peter Beaulieu, Puget Sound Council of Governments, Seattle, Washington, December 1, 1976.

70. Environmental Reporter, December 27, 1974, pp. 1328-1330.

71. Tri-State Regional Planning Commission, The Extent of Public Water and Sewer Systems in the Tri-State Region, New York, NY, May 1973.

72. Rivkin/Carson, Inc., Population Growth in Communities in Relation to Water Resources Policy, prepared for the National Water Commission, Arlington, VA, October 1971.

73. Hill, D., "A Growth Allocation Model for the Boston Region", Journal of the American Institute of Planners, 31(2), 1965.

74. Train, R.E., "The EPA Programs and Land Use Planning", Columbia Journal of Environmental Law, 2:255, Columbia University, New York, NY, 1976.

75. Breindenbach, A. and Strelow, R., EPA memorandum to Regional Administrators, Washington, DC, November 15, 1976.

76. Pailthorp, R., "Joint Treatment Rate Structures," CH2M Hill, Corvallis, OR, February 1977.

77. San Diego Building Contractors Association, Builder, 29(11): 13, November 1976.

78. Abt Associates, Manual for Evaluating Secondary Impacts of Wastewater Treatment Facilities, prepared for EPA Office of Air, Land and Water Use, Washington, DC, January 1976.

79. Booz-Allen & Hamilton, Inc., Methodologies for the Analysis of Secondary Air Quality Impacts of Wastewater Treatment Projects Located in AQMA, prepared for EPA Region II, New York, NY, March 1976.

80. Bureau of Regional Planning, Secondary Impact of Regional Sewerage Systems, State of New Jersey Department of Community Affairs, Trenton, NY, 1975.

81. Greenwood, E., "The Relationship of Science to the Practice Professions", Journal of the American Institute of Planners, 28:223-232, 1958.

82. Brodbeck, M., Readings in the Philosophy of the Social Sciences, The Macmillan Company, New York, NY, 1968.

83. Popper, K., The Logic of Scientific Discovery, Harper and Row Publishers, New York, NY, 1959.

84. Lowry, I., A Model of Metropolis, The RAND Corporation, Report No. RM-4035, RG, Santa Monica, CA, 1964.

85. Forrester, J., Urban Dynamics. The MIT Press, Cambridge, MA, 1969.

86. Seidman, D.R., The Construction of an Urban Growth Model, Delaware Valley Regional Planning Commission, Philadelphia, PA, 1969.

87. Center for Real Estate and Urban Economics, Jobs, People, and Land, Bay Area Simulation Study, University of California, Berkeley, CA, 1968.

88. Ohls, J.C. and Hutchinson, P., "Models in Urban Development" pp. 165-200 in: Saul I. Gass and Roger L. Sisson (eds), A Guide to Models in Governmental Planning and Operations, Sauger Books, Potomac, MD, 1975.

89. Blalock, H.M., Jr., Theory Construction, Prentice-Hall, Englewood Cliffs, NJ, 1969.

90. Meisel, W.S. and Collins, D.C., "Repro-Modeling: An Approach to Efficient Model Utilization and Interpretation", IEEE Transactions on Systems, Man and Cybernetics (1973) 3:349-358.

91. Personal communication, Mr. Robert McMahon, Old Colony Planning Council, Brockton, MA, November 16, 1976.

92. Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., and Bent, D.H. Statistical Package for the Social Sciences (2nd Edition), McGraw-Hill, New York, NY, 1975.

93. Andrews, D.F., "A Robust Method for Multiple Linear Regression," Technometrics, 16(4):523-531, November 1974.

94. McDonald. G.D. and Schwing, R.C., "Instabilities of Regression Estimates Relating Air Pollution to Mortality", Technometrics, 15(3):463-481, August 1973.

95. Hoerl, A.G. and Kernard, R.W., "Ridge Regression: Biased Estimation for Non-orthogonal Problems," Technometrics, 12:69-82, 1970.

96. Marquardt, D.W., "Generalized Inverses, Ridge Regression Biased Linear Estimation, and Nonlinear Estimation", Technometrics, 12:591-612, 1970.

97. Theil, H., "A Simple Modification of the Two-Stage Least-Squares Procedure for Undersized Samples", pp. 113-129, Arthur S. Goldberger and Otis D. Duncan (eds), Structural Equation Models in the Social Sciences, Seminar Press, New York, NY, 1973.

98. Office of Land Use Coordination, Mitigating Secondary Impacts from the Wastewater Facilities Program, U.S. Environmental Protection Agency, Washington, DC, 1977.

99. Abbett, R.W., American Civil Engineering Practice, John Wiley, New York, NY, 1956.

100. Metcalf & Eddy, Inc., Wastewater Engineering, McGraw-Hill, New York, NY, 1972.

101. Personal communication, Mr. Robert L. Michel, Municipal Construction Division, EPA, Washington, DC, November 1976.

102. Hudson, J.F., and Alford, M., Growth Effects of Major Land Use Projects: Wastewater Investments, Area of Analysis and Causal Model Specifications, Urban Systems Research & Engineering, Inc., Cambridge, MA, July 1977. This report was based in part on data from reference [63].

103. Great Lakes--Upper Mississippi River Board of State Sanitary Engineers, Recommended Standards for Sewage Works, Albany, NY, 1970.

104. Lexington-Fayette County Planning Commission, A Growing Community (1973 Update), Lexington, KY, June 1973.

105. Lerman, S., A Disaggregate Behavioral Model of Urban Mobility Decisions, Center for Transportation Studies Report No. 75-5, MIT, Cambridge, MA, 1975.

106. Peterson, G.E., Tax Policy and Land Conversion at the Urban Fringe, Land Use Center Working Paper No. 0875-04, Urban Institute, Washington, DC, 1974.

107. Environmental Impact Center, Secondary Effects of Public Investments in Highways and Sewers, Council on Environmental Quality, Washington, DC, 1974.

108. Personal communication, Mr. Michael Cook, Chief, Facility Requirements Branch, EPA, Washington, DC, August 26, 1977.

109. U.S. Urban Renewal Administration and Bureau of Public Roads, Standard Land Use Coding Manual, Washington, DC, January 1955.

110. Scott, G.R., "New Sewage Treatment Works for Boulder, CO, The American City, 74(1): 129-131, January, 1959.

111. Area Redevelopment Administration, Accelerated Public Works Program Directory of Approved Projects as of July 1, 1964, U.S. Dept. of Commerce, Washington, DC, August 1964.

112. Office of Water Programs, 1968 Inventory, Municipal Waste Facilities, Volumes I-X for each Federal region, EPA Publication No. OWP-1, Washington, DC, 1971.

113. Personal communicaction, Ms. Carol Wegrzynowicz, Municipal Construction Division, EPA, Washington, DC, June 21, 1977.

114. Facility Requirements Branch, Guidelines for 1976 Update of Needs for Municipal Wastewater Facilities, Environmental Protection Agency, Washington, DC, March 1976.

115. Evans, R., Cothran, C.L., and Cunningham, R.D., Multi-Model Transportation Site Selection Study, prepared for the Orange County Transportation District, CA and the City of Anaheim, CA, August 1976.

116. U.S. Departmentof Transportation, Guidelines for Trip Generation Analysis, Washington, DC, April 1973, p. 40.

117. Reference [44], pp. 30-37.

118. National Association of Regional Councils, Directory '77, Washington, DC, September 1976.

119. Chatterjee, and Cribbins, P.E., "Forecasting Travel on Regional Highway Network," Transportation Engineering Journal, Proceedings of the American Society of Civil Engineers, 98(TE2):210.

120. Maricopa Association of Governments, Trip Generation by Land Use, Prepared for the U.S. Department of Transportation, Federal Highway Administration, and Urban Mass Transit Administration, Phoenix, AZ, April 1975.

121. Tardiff, J., "Trip Generation as a Choice Process," Transportation Engineering Journal, 103(TE2):338-341, March 1977.

122. Gustafson, L., "An Empirical Study of Factors Influencing Trip Attraction and Trip Generation," High Speed Ground Transportation Journal, 7(3):314, 1973.

123. Meyerowitz, Trip Generation: Regression Analysis, Procedural Manual, Michigan Department of State Highways, Lansing, Michigan, May 1970.

124. Anderson, J.R., Hardy, E.E., Roach, J.T., and Witmer, R.E., A Land Use and Land Cover Classification System for Use with Remote Sensor Data, U.S. Geological Survey Paper #964, Washington, DC, 1976.

125. Harvard Institute of Economic Research, Time Series Processor, Harvard University, Cambridge, MA, December 1974.

126. Dixon, W.J., "Analysis of Extreme Values," Annals of Mathematics Statistics, 21:488-506, 1950.

127. Comsis Corporation, Travel Estimation Procedures for Quick Response to Urban Policy Issues, Volume II, prepared for Transportation Research Board, Washington, DC, February 1977.

128. Bates, J.W., Development and Testing of Synthetic Generation and Distribution Models for Urban Transportation Studies, State Highway Department of Georgia, Atlanta, GA, July 1971.

129. Jeffries, W. and Carter, E., Simplified Techniques for Developing Transportation Plans - Trip Generation in Small Urban Areas, Technical Bulletin #84, Engineering Experiment Station, West Virginia University, Morgantown, WV, December 1966.

130. Ashford, N. and Holloway, F., The Performance of Trip Generation Equations, UMTA Report No. URT-52(70)-71-2, Washington, DC, August 1971.

131. Asin, R.H., "Purposes for Automobile Trips and Travel", Nationwide Personal Transportation Survey Report #10, Federal Highway Administration, Washington, DC, May 1974.

132. Gibbs, L.L., Zimmer, C.E., Zoller, J.M., Source Inventory and Emission Factor Analysis: Volume I and II, EPA Publication No. EPA-450-3-75-082, Research Triangle Park, NC, 1974.

133. Turner, D.B., Workbook of Atmospheric Dispersion Estimates, EPA Publication No. AP-26, Research Triangle Park, NC, 1970.

# APPENDIX A
# PATH ANALYSIS

In this section we review briefly the concepts, assumptions, and problems associated with path analysis and its use in the GEMLUP-II study.

## 1. Aim of Path Analysis

Path analysis was developed by Sewall Wright [1-4] as a technique for studying and dealing with observed interrelated variables for which it can be assumed that there are several "ultimate" or exogenous variables that completely determine or cause them [5]. It is a method for studying the direct and indirect effects of variables taken as causes of other variables considered as effects [6]. It is not capable of deducing causal relations from available quantitative information (viz., correlation coefficients), but rather it is intended to combine this quantitative information with qualitative interpretation [1]. Thus, it is a technique useful in *testing* theories rather than in generating the, and it can be used to study the logical consequences of various hypotheses involving causal relations. In order to implement the technique, the researcher must make explicit a theoretical framework or model.

## 2. Path Diagrams

Figure A-1 presents various path diagrams which display graphically typical patterns of causal relations among sets of variables. Figure A-1a shows the relation between seven variables, $X_1$, $X_2$, $X_3$, $X_4$, U, V, and W, where $X_1$, $X_2$, $X_3$, $X_4$ are variables of the systems which are measurable. The variables U, V, and W are unmeasured and outside the system, but do affect variables in the system. These latter variables are often referred to as disturbances or residuals. Usually they are not represented in a path diagram and Figure A-1a and Figure A-1b can be considered the same path diagram. The residuals are "understood" to exist. The arrows in all the figures indicate the causal flow. Variable $X_1$ in Figure A-1a and A-1b is called an exogenous variable for its variability and is assumed to be determined by causes outside the system. Variables $X_2$, $X_3$, and $X_4$ of Figures A-1a and A-1b are called endogenous. Their variablility is explained by exogenous or endogenous variables in the system. Variables U, V and W of Figure A-1a can be considered exogenous.

Sometimes there is more than one exogenous variable in the system (see Figure A-1c and variables $X_1$ and $X_2$). In this case, the correlation between the two is represented by a curved line. The relation between the exogenous variables is usually not analyzed in path analysis.

The path diagram shown in Figure A-1d represents an interaction between two endogenous variables $X_2$ and $X_3$. Analysis of this type of system presents some serious technical problems which we will discuss later on*. This interaction is also termed a reciprocal causation or a feedback loop.

---

*See discussion of identification in Section 9 of this Appendix.

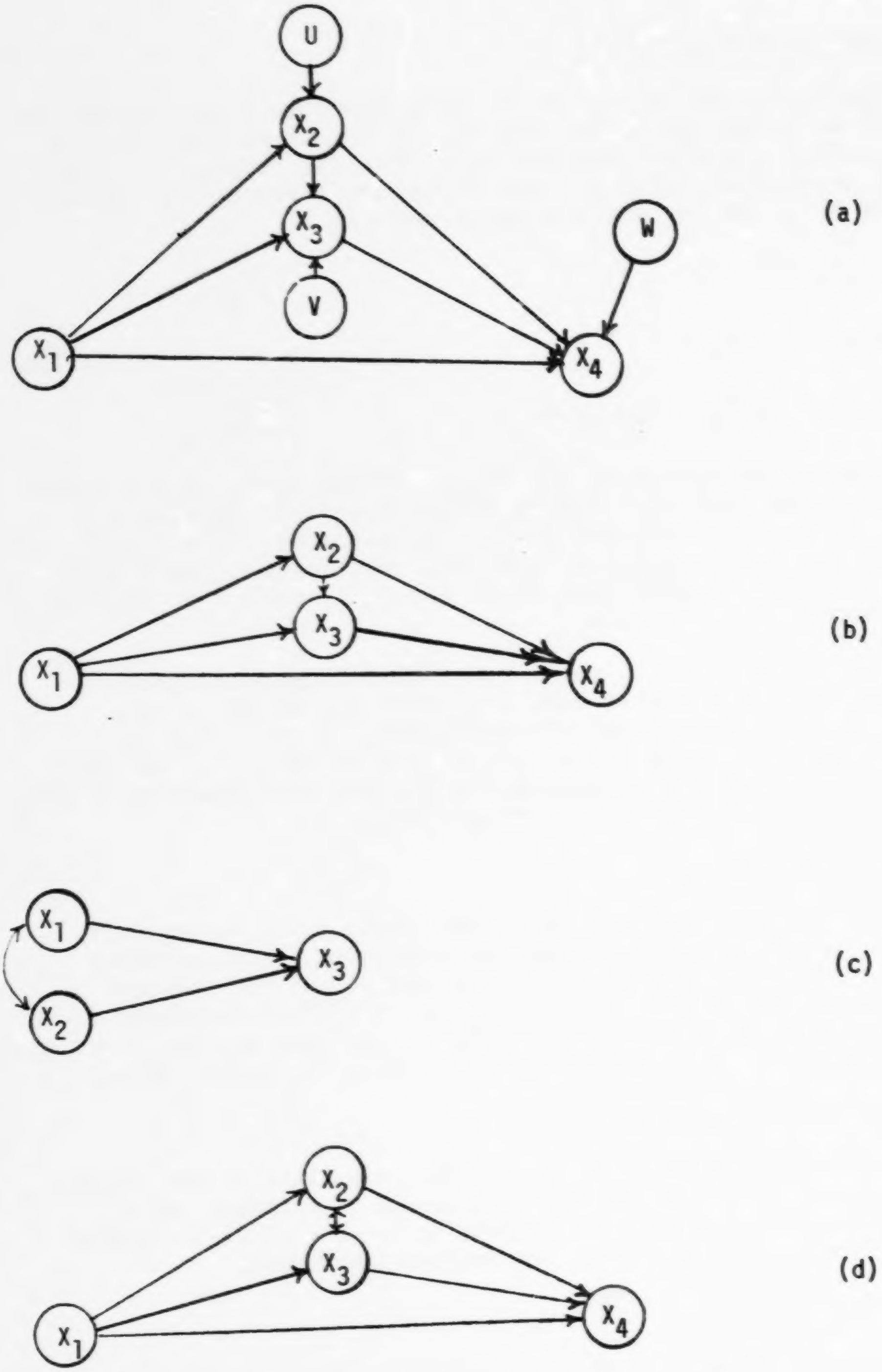

U
$X_2$
$X_3$
V
W
$X_1$
$X_4$
(a)
$X_2$
$X_3$
$X_1$
$X_4$
(b)
$X_1$
$X_3$
$X_2$
(c)
$X_2$
$X_3$
$X_1$
$X_4$
(d)

FIGURE A-1 ILLUSTRATIVE PATH DIAGRAMS

### 3. Approach (Structural Model)

In order to analyze the system of variables under investigation the path diagram must be formalized mathematically. A structural model must be developed which defines a causal structure specifying the network of causal paths that exist between variables and identifying the parameters of causation. For Figure A-la, the mathematical model would be:

$$\begin{aligned} X_2 &= b_{21}X_1 + U \\ X_3 &= b_{31}X_1 + b_{32}X_2 + V \\ X_4 &= b_{41}X_1 + b_{42}X_2 + b_{43}X_3 + W \end{aligned} \qquad \text{(A-1)}$$

These equations are called structural equations. If the variables are standardized (i.e., have mean zero and standard deviation unity), then the b's are called path coefficients. If they are not standardized, the b's are called path regression coefficients or structural coefficients (there would then also be an intercept in each equation of (A-1)). The b's are the parameters quantifying the causal effects.

Given a structural model, one has an explicit and quantitative statement of theory and so it can be used to explain why variables vary together as they do. It is possible with such a model to calculate how a change in any one variable in the system will affect the values of the other variables. Also, it is possible to analyze how changing the structure of the system would affect the character of the system [7].

### 4. Connection Regression

It should be emphasized that the three equations of system (A-1) each separately represent a multiple regression equation. The estimates of the b's are obtained using ordinary least squares (i.e., the standard regression technique). Up to this point, there is no difference between path analysis and standard multiple regression. The b's are beta weights if the X's are standardized, otherwise they are partial regression coefficients.

### 5. Assumptions and Deviations

The transition from the path diagram of Figure A-la to the system of structural equations (A-1) and the appropriateness of standard least squares (i.e., multiple regression) for estimating the b's of (A-1) involve a number of very important assumptions. We now consider these.

#### a. Linearity

The change in one variable is assumed to always occur as a linear function of changes in other variables. If it is believed or known that the relation is nonlinear, for example, say multiplicative:

$$X_3 = e^{(b_{31}X_1 + b_{32}X_2 + U)} \qquad \text{(A-2)}$$

then the data must be transformed before the analysis. If there is non-linear "causality" the usual path analysis will not detect it.

### b. Recursive

In order for the above discussion to be correct, the system must not contain reciprocal causation or feedback loops (e.g., as in Figure A-1d with the variables $X_2$ and $X_3$). The causal flow must be unidirectional. At a given point in time, a variable cannot be both a cause and an effect of another variable. Such a system is called recursive.

It is possible to solve a system of structural equations if there are feedback loops by use of instrumental variables. Instrumental variables are exogenous variables introduced into the system for this specific purpose. They are used routinely in economic analysis [8,9]. Basically what they do is to put enough constraints on the structural system so that by use of estimation techniques (either two-stage least squares or limited information least-generalized residual-variance solution) [8] the path coefficients can be estimated. A point often not emphasized in the "traditional" sources of path analysis is that little is known about the resulting estimates of the path coefficients. The most one can say is that they are consistent estimates of the true b's. This means that they converge in probability to the true b's or, in other words, as the sample size increases they get "closer" to the true b's. Basically, nothing is known about the estimates for small samples (such as 40 in the current study). All the "nice" properties of the multiple regression estimates (i.e., minimum variance, unbiased estimates) do not hold for the new ones.

In this study we made use of the above techniques (i.e., use of instrumental variables and two stage least squares) to solve for the b's. In order to judge the usefulness and appropriateness of the estimated b's, four different approaches were tried. The first three involved the use of approximate t tests (to judge if the b's are statistically significantly different from zero), beta weights (to judge if in the standardized scale the magnitude of the effect is "large") and the signs of the b's (to verify that the signs of the b's are what theory would predict). The fourth approach involved the use of a technique called "Jackknifing"* to verify that the b's were stable (i.e., do not change radically with the inclusion or exclusion of just one observation).

Another way of solving the system if feedback loops exist is to use lead-lag relations. This might be possible in longitudinal data [4]. Consider the path diagrams of Figure A-2. In Figure A-2a, we have an interaction between variables $X_2$ and $X_3$. However, say $X_2$ and $X_3$ represent the final states after a time period (say, 2 hours, for example). Further, say it is possible to break the 2 hours into two time periods (time periods 1 and 2) (see Figure A-2b). In time period 1, we may have variable $X_2$ (called $X_{21}$) as a cause of $X_3$ (called $X_{31}$) but not $X_{31}$ as a cause of $X_{21}$. Then into time period 2 we have $X_3$ as a cause of $X_2$ (i.e., $X_{31}$ $X_{22}$). The system shown in Figure A-2b is recursive and given sufficient data it is solvable. The dynamics of the system over the full time period are needed. Unfortunately, for the present study, sufficient data were not available to allow us to implement this lead-lag device.

---

*The concept of "Jackknifing" is discussed further in Section 5 of this Appendix.

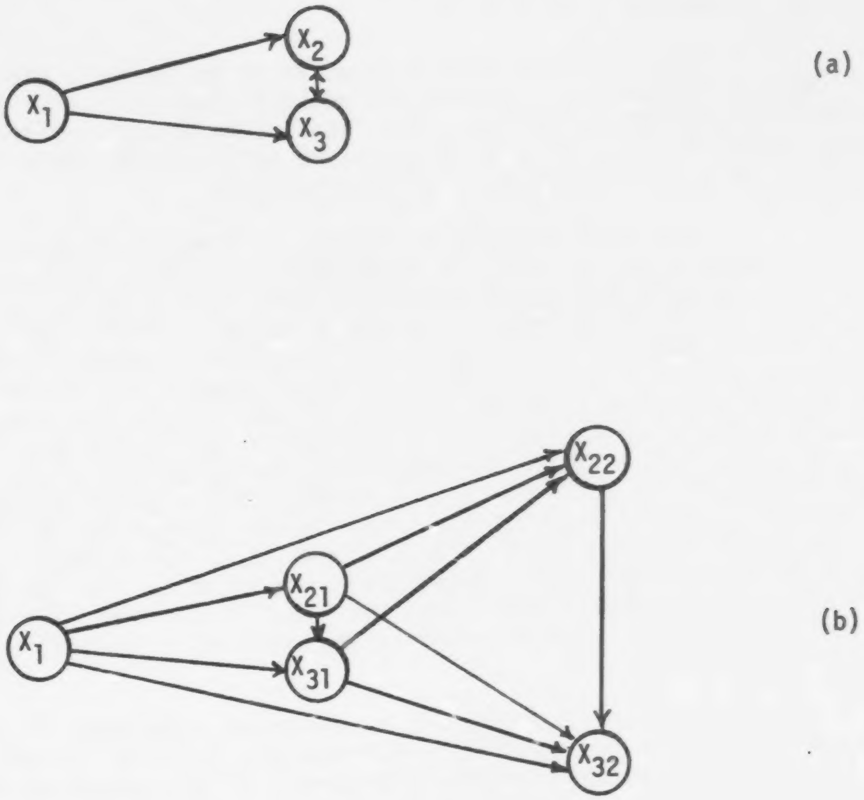

X1
X2
X3
(a)
X1
X21
X31
X22
X32
(b)

FIGURE A-2 LEAD - LAG RELATIONSHIPS

### c. Causal Priorities

The causal priorities among the variables must be "correctly" established outside of the analysis. A major drawback of path analysis is that it cannot establish causal relationships using data alone. One needs as input both data and theory. The ordering of the variables (that is, the direction of the arrows in the path diagrams) must be established by the investigator. "There is no error-check mechanism in path analysis to reject an incorrect ordering" [7]. It is common that the data can fit well two very different orderings. The selection of the appropriate one depends on correct theory. When one is dealing with variables ordered in time it is often possible to know the causal priorities. Path analysis seems well suited for this type of problem. For the present study, the ordering in time was not available, and so a substantial effort was required to establish causal priorities. Also, causal priorities were established by use of instrumental variables and two stage least squares for estimating the b's in feedback loops (e.g., for the cases where it is not clear that a causal relationship exists, if the estimated path coefficients (b) is zero then the analysis indicates there is no causal relationship).

### d. Uncorrelated Residuals of Endogenous Variables

In the actual analysis the observed correlations among the observed variables are used to estimate or identify the path coefficients. If the residual (i.e., the disturbances) are uncorrelated, then the usual least squares technique (i.e., multiple regression) can be used for the estimation. If, however, the residuals are correlated, then the path coefficient cannot be identified by the ordinary least squares technique for there are then too many unknowns. An implication of this assumption of uncorrelated residuals is that all inputs must be entered explicitly into the analysis. (It is possible to ignore a variable that affects two or more inputs as long as it does not affect the residuals of any endogenous variables). This restriction of uncorrelated residuals is very important and meaningless and confusing results may come out of the analysis if it is not met.

Because of its importance to the project under consideration, we shall develop a bit further here the consequences of not including all inputs explicitly in the model. Say we have three variables $X_1$, $X_2$, and $X_3$ which are causal factors of a dependent variable Y. The appropriate equation is:

$$Y = b_{y1}X_1 + b_{y2}X_2 + b_{y3}X_3 + e \qquad \text{(A-3)}$$

Here, e is the residual. Now say we do not include $X_3$ in our system and we "believe" the correct equation is:

$$Y = b_{y1}X_1 + b_{y2}X_2 + e' \qquad \text{(A-4)}$$

e' in (A-4) is a new residual. It is actually $e+b_{y3}X_3$. It can be shown [8] that the estimate we will obtain for $b_{y1}$ will actually be an estimate not of $b_{y1}$ alone, but rather of

$$b_{y1} + b_{y3}\,(B_{31.2}) \qquad \text{(A-5)}$$

Where $b_{y3}$ is the real path coefficient given in (A-3) and $B_{31.2}$ is the regression coefficient of $X_3$ and $X_1$ after allowing for the effects of $X_2$. *The important point here is that if we leave out an input it will reappear in our path coefficients.* We will have no way of knowing how much of our estimate is the real $b_{y1}$ and how much is $b_{y3}(B_{31.2})$. It is possible that $b_{y1}$ is positive yet the value of (A-5) is negative (a complete change in sign). This point is not made very clear in much of the existing literature on path analysis.

Another implication of this fourth assumption is that all residuals should be uncorrelated with all other exogenous variables that precede them in the system. If one suspects that the residuals are correlated, then the technique of two stage least squares can be used to remove this problem. However, in doing so, it introduces the previously discussed problem of having estimates whose statistical properties are not well known. In the present project, for those equations not requiring two stage least squares estimation techniques, we used a sufficient number of relevant variables in the model to minimize this problem of correlated residuals.

### e. Stability of the Estimates

In order for path analysis to be of use the resulting estimates should be stable. They will be of little use if they change radically. depending on the sample observations. In most applications of path analysis this is achieved by using large samples and reliable measuring instruments. Tukey [10] and Turner and Stevens [11] with regard to this point suggested using path regression coefficients rather than the usual path coefficients (i.e., beta weights). Path regression coefficients tend to be more stable as one applies the same structural model to different populations.

For the current project we had only 40 observations for use in estimating model parameters. This is a small sample, and so there may be a problem of obtaining stable estimates. One way of judging the stability of the estimates obtained is to use the statistical technique of "Jackknifing". Instead of running one regression using 40 observations, Jackknifing involves performing 40 regressions each containing 39 observations and then examining the stability of the sets of path coefficients obtained. This technique prevents the path analysis from capitalizing on chance elements in the data. See Section IV for a full discussion of the application of this study.

### f. Multicollinearity

If the correlations between independent variables in a multiple regression are extremely large in absolute value, we have what is called the problem of multicollinearity. If present, it is usually impossible to separate the influences of the variables involved. The effect on the path analysis will be to produce unstable path coefficients (they will have large standard errors) and beta weights greater than unity in value. Obviously, the purposes of path analysis can be defeated if collinear variables are present. Notice these collinear variables can be both exogenous and endogenous as long as they appear somewhere as independent variables.

To handle the problem of multicollinearity we examined the zero order (i.e., simple) correlations between variables for large correlations and paid careful attention to the output of individual regressions looking for unusually large (in absolute value) beta weights and standards errors for the regression coefficients.

g. Suppressor Variable Problem

It is possible for a path analysis system to contain too many input variables. For example, say $X_1$ refers to the size (diameter) of the main interceptor line and $X_4$ is commercial land use in the area of interest. Further, say $X_2$ and $X_3$ are two other variables, the structural model is given by the equations,

$$\begin{aligned} X_2 &= b_{21}X_1 + b_{2u}U \\ X_3 &= b_{31}X_1 + b_{3v}V \\ X_4 &= b_{41}X_1 + b_{42}X_2 + b_{43}X_3 + b_{4w}W \end{aligned} \qquad \text{(A-6)}$$

and the simple correlation matrix is:

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| $x_1$ | 1.000 | .800 | .600 | .000 |
| $x_2$ | | 1.000 | .480 | .288 |
| $x_3$ | | | 1.000 | .384 |
| $x_4$ | | | | 1.000 |

Given this it can be shown that the solution for $X_4$ is:

$$X_4 = -X_1 + .8X_2 + .6X_3 + .736W \qquad \text{(A-7)}$$

The size of the sewer line appears to have a negative impact on commercial land use (path coefficient is -1).

The actual situation is that a model was developed to produce the correlation matrix and it was:

$$X_4 = .48U + .48V + .736W \qquad \text{(A-8)}$$

U,V and W are the residuals of $X_2$,$X_3$ and $X_4$. $X_4$ and $X_1$ are actually uncorrelated. The "strange" result shown in (A-7) arises because of the following: If we use $X_2$ and $X_3$ to predict $X_4$ we will have a useful result because $X_2$ and $X_3$ contain U and V which are related to $X_4$. However, $X_2$ and $X_3$ also contain $X_1$ which is irrelevant. In (A-7) the term $X_1$ is used to "suppress" the irrelevant components in $X_2$ and $X_3$. That is, $X_2$ and $X_3$ do contain a portion of $X_1$ in them and this portion is subtracted again in the equation by adding the negative term with $X_1$. Figure A-3 illustrates the problem. The system of equation (A-6) is assuming Figure A-3a while the true situation is Figure A-3b. The importance of this illustration is that one cannot just include anything into a structural system and believe all will work out well.

To handle this problem we examined the zero order correlations between all endogenous variables (i.e., those that are dependent variables in some regressions) with all other variables. Those cases where zero or near zero correlations appeared were noted. Later, if the partial relationship proved to be significant, an examination was performed to see if a suppressor variable problem existed by investigating the stability of the model coefficients including/excluding the variable.

### 6. Other Assumptions

There are other assumptions that deal solely with the fact that path analysis is a regression technique (e.g., independence of observations). We will not enumerate them for they are standard [12].

### 7. Theory Trimming

Often the approach used in path analysis is to develop the most elaborate system and then, after estimating the involved path coefficients, attempt to refine or trim the system by dropping those paths that have coefficients close to zero. Those path coefficients near zero imply no or only a weak causal linkage between the involved variables. The possibility for refining or trimming a theory and thus making a theory more parsimonious is of considerable significance.

There are a number of ways of deciding which path coefficients can be dropped. First, there is the criterion of statistical significance e.g., drop all path coefficients which do not differ significantly in the statistical sense (0.05 level of significance) from zero. With large samples, this could imply retaining some very small coefficients and with small samples this may imply dropping some very large coefficients. Second, there is the possibility of dropping all coefficients whose values are less than some preassigned fixed value (e.g., say 0.05 or 0.03) [13]. There is also the dependence analysis procedure of Boudon [14]. In this technique, one usually assumes that some path coefficients of the structural equations are zero. To test this procedure, one sees how well the non-zero paths reproduce the original correlation matrix. In general, for any trimming procedure it is important to test how deletion of some path coefficients affects the ability to reproduce the original observed correlations (an important property of path analysis itself is the ability of reproducing the original observed correlations by means of the path coefficients).

The approach used in trimming the model is outlined in Section IV and consisted of developing a comprehensive set of statistical test criteria.

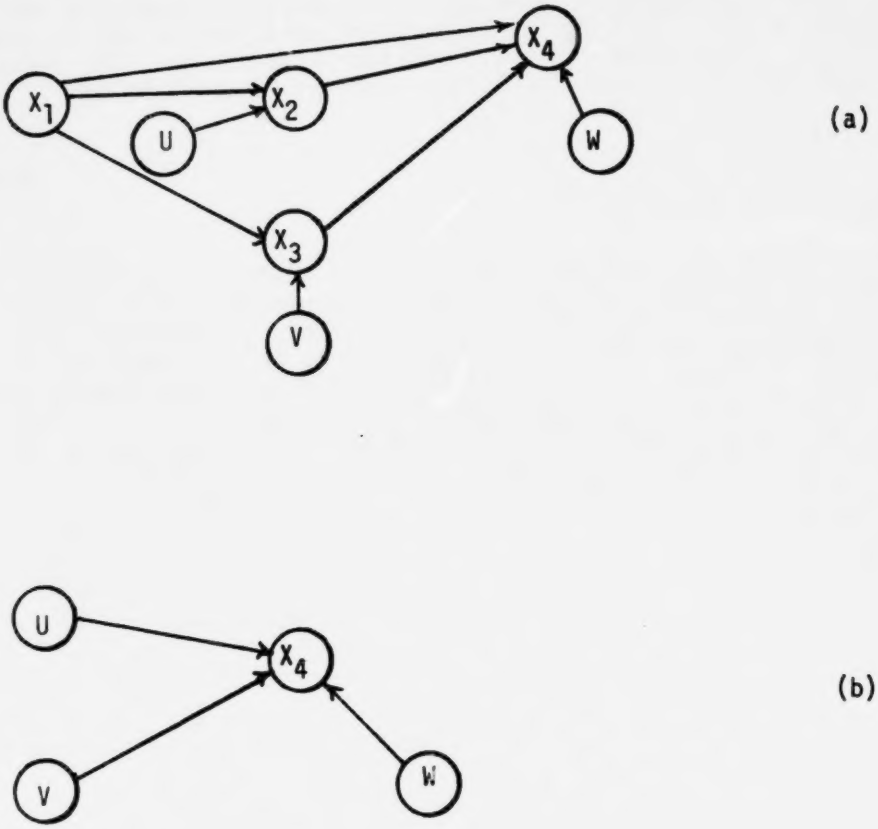

X1
X2
X3
X4
U
V
W
(a)
U
X4
V
W
(b)

FIGURE A-3 ILLUSTRATION OF SUPPRESSOR VARIABLE PROBLEM

## 8. Quantification of Direct and Indirect Effects

An important application of path analysis is the analysis or decomposition of a correlation into its various components [6]. For a given causal model it is possible to determine what part of a correlation between two variables is due to the direct causal effects and what part is due to the indirect effects. As an example of this consider Figure A-4a. The system of structural equations that correspond to the model of Figure A-4a is:

$$X_2 = b_{21} X_1 + U \qquad \text{(A-8)}$$

$$X_3 = b_{31} X_1 + b_{32} X_2 + V$$

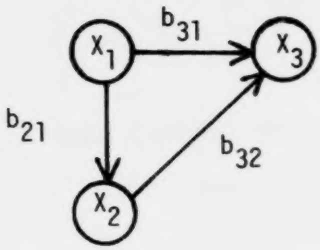
For the sake of simplification let us assume the X's are standardized (i.e., have mean zero and unit standard deviations). Suppose we wish to analyze the correlation between the variables $X_1$ and $X_3$. Let $r_{31}$ represent this correlation. Now there are two ways that $X_1$ affects or influence $X_3$. First, there is the direct effect of $X_1$ and $X_3$, represented in Figure A-4a by the arrow going directly from the variable $X_1$ to the variable $X_3$. This effect is quantified by the path coefficient ($b_{31}$) in the second equation of (A-8); that is, a one unit change in $X_1$ produces a ($b_{31}$) unit change in $X_3$. Second, $X_1$ effects $X_2$, which in turn effects $X_3$. This indicates that $X_1$ directly effects $X_3$ through $X_2$. By use of standard manipulations it can be shown that this indirect effect of $X_1$ and $X_3$ is quantified by the product ($b_{32}$ $b_{21}$). This represents the effect of $X_1$ on $X_2$ ($b_{21}$) and then the effect of $X_2$ on $X_3$ ($b_{32}$). Further manipulation can be done to show that the correlation $r_{13}$ is equal to:

$$r_{13} = b_{31} + b_{32} b_{21} \qquad \text{(A-9)}$$

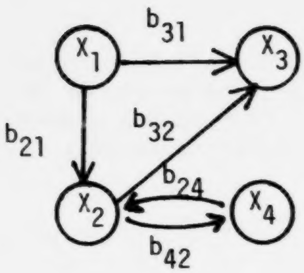
That is, the *net* effect of multiple direct and indirect relationships is the *sum* of the individual effects.

The above ideas can be extended so that path diagrams can be "read" to define the net effect from one variable to another, even when the variables are connected in complex systems involving multiple chains of causation and intervening feedback loops. Such analyses allow one to trace the reverberations of model inputs; providing insight into how system variables will be correlated in empirical data, and how changes in inputs would effect the system's endogenous variables. Figure A-4b offers a more complex path analytic model, including a feedback loop between variables $X_2$ and $X_4$. Suppose we wish to analyze the total effect of $X_1$ on $X_3$. As before, there is a direct effect ($b_{31}$), but there are now numerous indirect effects. A one unit change in $X_1$ produces a change of ($b_{21}$) units in $X_2$. This causes a change of ($b_{21}$ $b_{42}$) units in $X_4$. The change in $X_4$ in turn causes another change in $X_2$: ($b_{21}$ $b_{42}$ $b_{24}$). So after one cycle, $X_2$ is incremented by the value of a return effect in the feedback loop. This effect can in turn move around the loop again, causing still another change in $X_2$ of ($b_{21}$ $b_{42}^2$ $b_{24}^2$) units, and that in turn can move around the loop once more, ad infinitum. The *net* effect of $X_1$ on $X_2$ is therefore:

$$(b_{21}) + (b_{21} b_{42} b_{24}) + (b_{21} b_{42}^2 b_{24}^2) + \ldots (b_{21} b_{42}^n b_{24}^n) + \ldots$$


$X_1$
$X_2$
$X_3$
$b_{31}$
$b_{21}$
$b_{32}$

(a) Simple Case


$X_1$
$X_2$
$X_3$
$X_4$
$b_{31}$
$b_{21}$
$b_{32}$
$b_{24}$
$b_{42}$

(b) Complex Case

FIGURE A-4 ILLUSTRATION OF THE DIRECT AND INDIRECT CAUSAL RELATIONSHIPS IN PATH DIAGRAMS

This infinite geometric series can be expressed more simply by the expression:

$$\frac{b_{21}}{1-L}$$

where

$L = (b_{42}\ b_{24})$, the "return effect" of the feedback loop.

The total indirect effect of $X_1$ on $X_3$ is therefore just:

$$\frac{(b_{21}\ b_{32})}{1 - L}$$

and the net effect of $X_1$ on $X_3$ is:

$$b_{31} + \frac{(b_{21}\ b_{32})}{1 - L}$$

The determination of the total effect of variables on one another in a complex path analytic system is, in general, not as simple as the preceeding example would indicate. Heise [15] has developed a generalized approach to determining such effects, and the reader is referred to that text for a full presentation of the techniques involved.

In the current project, we performed a decomposition analysis of correlations into direct and indirect effects for all significant variables contained in the final trimmed path analysis model.

### 9. Identification

In this final section we briefly review the limitations imposed on the model specification due to the statistical requirement of *identification*.

When there are feedback loops or reciprocal causal paths in a path analytic model, then the residuals in the model are correlated with the endogenous variables and use of ordinary least square techniques will yield biased and inconsistent estimates of the structural coefficients. For example, say $Y_1$ and $Y_2$ are two land use variables measured in appropriate units and say there is believed to be a feedback loop between these. Then possible structural equations are:

$$Y_1 = a_1 + b_{21}Y_2 + e_1 \qquad \text{(A-10)}$$

$$Y_2 = a_2 + b_{21}Y_1 + e_2$$

It can be shown [16] that the residual $e_1$ is correlated with $Y_2$ (since $Y_2$ is a function of $Y_1$) and the residual $e_2$ is correlated with $Y_1$. Ordinary least squares (which assumes *un*correlated residuals) applied to either of these equations will result in biased and inconsistent estimates of the structural parameters $a_1$, $b_{21}$, $a_2$, and $b_{12}$. By biased and inconsistent estimates we mean that the average value of the estimates will *not* equal the values of the structural parameters no matter how large the sample size may be.

The approach usually followed at this point to obtain appropriate estimates of the structural parameters is to introduce instrumental variables and apply two stage least squares. This approach will yield consistent estimates if the relevant equations are properly identified. The term "instrumental variable" merely means a variable which acts as an instrument in making the system identifiable. Below, we review how instrumental variables are introduced, how two stage least squares is applied and how the identifiability condition arises.

Given the structural system of equation A-10, one must introduce at least one new variable which is a "cause" of $Y_1$ but not for $Y_2$. Let us call this variable $X_3$. Similarly an $X_4$ should be introduced for $Y_2$. $X_3$ and $X_4$ are instrumental variables. There is no reason why there need be only one instrumental variable for each endogenous variable, however there must be *at least one*. To allow for more generality let us say we have a second instrumental variable for $Y_1$ and let us call it $X_5$. The important point is that we need some instrumental variables which are causally related to one of the endogenous variables, but not to the other in the system. It is possible to have some exogenous variables (which are still often called instrumental variables) which are causes of more than one of the endogenous variables. For our example, let us say $X_6$ is such a variable.

The revised structural system is now,

$$\begin{aligned} Y_1 &= a_1 + b_{21}Y_2 + b_{31}X_3 + b_{51}X_5 + b_{61}X_6 + e'_1 \\ Y_2 &= a_2 + b_{12}Y_1 + b_{42}X_4 + b_{62}X_6 + e'_2 \end{aligned} \qquad \text{(A-11)}$$

where $e'_1$ and $e'_2$ are not the same residuals as $e_1$ and $e_2$ of (A-10).

It can still be shown that $Y_2$ is correlated with $e'_1$ and $Y_1$ is correlated with $e'_2$ (i.e., we have correlated residuals), so ordinary least squares still cannot be applied directly to obtain consistent estimates of the structural parameter b's. However, if we manipulate the equations of (A-11), a new system of equations which have $Y_1$ and $Y_2$ only on the left, viz.,

$$\begin{aligned} Y_1 &= a'_1 + b'_{31}X_3 + b'_{41}X_4 + b'_{51}X_5 + b'_{61}X_6 + e''_1 \\ Y_2 &= a'_2 + b'_{32}X_3 + b'_{42}X_4 + b'_{52}X_5 + b'_{62}X_6 + e''_2 \end{aligned} \qquad \text{(A-12)}$$

is possible. This new set of equations gives us a way of estimating the b's of (A-11). Notice these equations have all the instrumental variables on the right hand side. The equations of (A-12) are often called the *reduced form equations*. Ordinary least squares may be applied to obtain estimates of the a's and b's (and in turn estimates of $Y_1$ and $Y_2$ are possible). This is the first stage in two stage least squares analysis.

Next, the estimates of $Y_1$ and $Y_2$ obtained from (A-12) can be substituted into the right hand side of the equations in (A-11). Using $\hat{Y}_1$ and $\hat{Y}_2$ to represent these estimates, the system (A-11) with this adjustment becomes:

$$\begin{aligned} Y_1 &= a_1 + b_{21}\hat{Y}_2 + b_{31}X_3 + b_{51}X_5 + b_{61}X_6 + e'_1 \\ Y_2 &= a_2 + b_{12}\hat{Y}_1 + b_{42}X_4 + b_{62}X_6 + e'_2 \end{aligned} \qquad \text{(A-13)}$$

$\hat{Y}_2$ is *not* correlated with $e'_1$, since $\hat{Y}_2$ is a function of only the exogenous variables $X_3$, $X_4$, $X_5$ and $X_6$ and not $Y_1$, and for similar reasons $\hat{Y}_1$ is *not* correlated with $e'_2$. Thus, ordinary least squares may now be applied to obtain consistent estimates of structural parameters, the b's. This is the second stage in two stage least squares analysis.

In order to obtain consistent estimates of the b's it is necessary that $\hat{Y}_2$ is not a linear function of $X_3$, $X_5$ and $X_6$ alone. This is the identification problem (we want to "identify" or estimate the b's) and by bringing $X_4$ in the system we have this necessary condition satisfied. Recall $X_4$ is the instrumental variable that is causally related to $Y_2$ but not $Y_1$. A similar necessary condition is needed to identify the b's of the second equation of (A-13).

The necessary condition for identification can be formalized as follows: suppose we want to estimate the b's of the first equation of the system (A-11). First, one counts the number of instrumental variables in the full system. Here, there are 4 ($X_3$, $X_5$, $X_6$ and $X_4$). Next, one counts the number of instrumental variables excluded from the equation under consideration. Let us call this number $m_0$. Here $m_0$ is 1 for variable $X_4$. Then we count the number of endogenous variables on the right hand side of the equation. Let us call this number q: q for this problem is 1 for $Y_2$. The *necessary* condition for identifying the parameters of the equation under consideration is then [16]:

$$m_0 \geq q - 1 \qquad \text{(A-14)}$$

For our example, $1 \geq 1 - 1 = 0$ so we have identification. Condition (A-14) is necessary but not *sufficient* (i.e., if an equation is identified it must satisfy (A-14), however, there are equations that satisfy (A-14) but are not identified). The sufficient condition involves rank conditions of various matrices [16]. Because we anticipated the existence of feedback loops in the GEMLUP-II model, the problem of identification required careful consideration. In the model specification phase of this project, care was taken to specify exogenous variables that could serve as instrumental variables. Additionally, the necessary conditions for identification (A-14) was continually checked for compliance.

## B. CAUSAL ANALYSIS IN GEMLUP PROJECTS*

This chapter attempts to explain why a causal analysis of secondary growth effects is undertaken in GEMLUP projects. After the GEMLUP I reports [17-19] were distributed the need for such an explanation became obvious because the Project Officer received a number of questions relating to the issue of causality. The questions went something like this: "How do you know that the land uses you are predicting are really secondary development induced by the major project?" This chapter is a long answer to that short question. There are many subtle issues involved, and a large body of literature exists concerning each issue. The discussion here addresses only the main issues in as straightforward a way as possible.

### 1. Prediction, Explanation, Theory, and Model

GEMLUP projects contain tasks that directly involve prediction, explanation, and modeling [20,21]. The projects indirectly involve theory formulation and testing. The overall purpose of the projects is to develop prediction equations to estimate the amount and type of secondary land use induced by constructing and operating a major project, such as a wastewater facility, a large residential complex, or an office/industrial park. To ensure that the prediction equations are formulated on a firm scientific basis, a causal analysis of secondary development is undertaken to explain the structural relations that we think exist in induced development. The explanation takes the form of setting up and testing a model of a theory of induced development. The model is made up of the path analytic diagrams and equations described in the first volume of each GEMLUP final report, and it is a mathematical description of the verbal theory hypothesized in the same volume. All of these concepts are elaborated upon below.

#### a. Prediction and Explanation

The essence of science is to explain. Usually the focus is to explain empirically defined phenomena according to commonly accepted scientific (definite) criteria [22]**. Explanation is formally defined as a process by which singular events are related to other events through the use of appropriate

---

* This chapter was written by the EPA Project Officer, Thomas McCurdy.

**Criteria involve: A definition of what is acceptable evidence, how the conduct of scientific inquiry should proceed, the capability to generalize from the evidence, and the minimum standards for internal and external validity [22]. These issues are the province of "philosophy of science."

**Preceding page blank**

general statements or theories [22].+ Singular events and isolated phenomena have no meaning: they are only facts and have no significance without generalization into widely applicable laws governing the behavior of empirical events or objects. It is these laws and generalizations that allow us to predict the outcome of unknown events [22].

Saying the same thing less formally, we have to be concerned with generalization and explanation in GEMLUP so that we can predict. While explanation and prediction are not identical, they are obviously related. Kaplan in the The Conduct of Inquiry says that from the viewpoint of a philosopher of science, the ideal explanation is one that allows prediction [23:349-350].

> Predictions can be and often are made even though we are not in a position to explain what is being predicted. This capacity is characteristic of well-established empirical generalizations that have not yet been transformed into theoretical laws . . . In short, explanations provide understanding, but we can predict without being able to understand and we can understand without necessarily being able to predict. It remains true that if we can predict successfully on the basis of certain explanations we have good reason, and perhaps the best sort of reason, for accepting the explanation.

Kerlinger and Pedhazur [6] bring the subject closer to home in their discussion of the role of regression analyses in explaining and predicting phenomena. This discussion is particularly apropos here because GEMLUP projects use multiple regression analysis for both purposes, albeit in slightly different forms. Stepwise multiple regression analysis is used for purely predictive work in GEMLUP, while path analysis, which is actually multiple regression analysis used in a particular way, is used for explanatory work. Computationally, the techniques are very closely related.

> Regression analysis can play an important role in predictive and explanatory research frameworks. Prediction and explanation reflect different research concerns and emphases. In prediction studies the main emphasis is on practical application. On the basis of knowledge of one or more independent variables, the researcher wishes to develop a regression equation to be used for the prediction of a dependent variable, usually some criterion of performance or accomplishment. The choice of independent variables in the predictive framework is determined primarily by their potential effectiveness in enhancing the prediction of the criterion.

---

+There is a hierarchy of explanation organized according to how widely applicable are the findings and results of scientific endeavor. The hierarchy follows, ranked from the least to the most generally applicable [22,24].

1. Strict definition and description of facts, or observations.
2. Order and classification of facts in a typology.
3. Comparison and relation of facts according to a model.
4. Generalization of facts into a theory.
5. Generalization of facts into an empirical law, or a universal generalization.
6. Generalization of facts into a fundamental generalization.
7. Generalization of facts into an axiom.

The causal analyses in GEMLUP projects involve generalizations of the type listed in both items three and four, as they involve developing and testing a "factor theory," which is a limited generalization. This will be discussed in more detail later.

In an explanatory framework, on the other hand, the basic emphasis is on the explanation of the variability of a dependent variable by using information from one or more independent variables. The choice of independent variables is determined by theoretical formulations and considerations. Stated differently, when the concern is explanation, the emphasis is on formulation and testing explanatory models or schemes. It is within this context that questions about the relative importance of independent variables become particularly meaningful. Explanatory schemes may, under certain circumstances, be enhanced by inferences about causal relations among the variables under study [6].

A concrete but simplified example would be that if a researcher is interested only in prediction, then the focus is on the magnitude of $R^2$, the coefficient of determination, and its statistical significance, but if the research is interested in explanation, the *relations* between the dependent variable and its set of independent variables are important [6]. The focus is thus on the entire regression equation and the regression coefficients, the $\beta$'s (if standardized; b's if not standardized). Squared semi-partial correlations,* in particular, are of prime interest because they identify the contribution to the variance of the dependent variable that each independent variable adds *after* the variance contribution of preceding variables are accounted for [6]. This is statistical control of independent variable effects, analogous to control of independent variable impacts in a classic experimental design. It is the conceptual basis for the causal analytic work done in GEMLUP, which allows us to infer the induced secondary developmental impacts of major land use projects. It allows us to explain secondary development caused by large projects.+

### b. Theory and Model

As discussed above, a theory is a specific type of generalization. There are a number of definitions of the word, but one good one is "a plausible or scientifically acceptable general principle or body of principles offered to explain phenomena" [30]. An alternative definition is a deductively related group of general statements that explain observable phenomena [22]. A theory is an abstract and symbolic construction that relates general statements on the basis of underlying similarities; it is not an inductive generalization from a body of observations.

---

*A semi-partial correlation represents the correlation between two variables with the influence of another variable or other variables removed from one of the variables being correlated. The square of the semi-partial correlation coefficient indicates the amount of variance contributed by separate independent variables in an regression equation, as mentioned above in the text. The variance contributed, however, is positionally dependent because it is the variance contributed after that denoted by the preceding variables [6].

An example symbol for a semi-partial correlation is $r_{y(2.1)}$, which means the correlation between the dependent variable y and variable 2 controlled for, or partialled from, the correlation between variables 1 and 2.

For an extended discussion of partial and semi-partial correlation, see pp. 84-99 in Kerlinger and Pedhazur [6].

+This point is elaborated upon in the section on causal analysis.

While the discussion of theory so far seems very academic, GEMLUP projects indirectly involve theory testing and formulation. The kind of theory investigated in GEMLUP, however, is a special type of theory, called factor theory. It is the most rigorous type of theory that we could ascribe to.*

Factor theory is characterized by narrow and non-overlapping generalizations, a selective, explicit enumeration of all factors hypothesized to influence a given phenomen, and the use of empirically defined variables to represent factors involved in the theory [25]. The most theoretical aspects of this type of theory lie in the selection of the general statements in the theory and the positing of rules of interaction governing their application.

Almost every effort at causal explanation involves factor theory because of its explicit enumeration of (all) factors leading to a particular development. However, it is limited theory because it does not readily suggest other generalizations [22]. The independent selection and arrangement of general statements focused on a definite phenomenon in factor theory precludes such generalizations.

A model, on the other hand, depicts the structure of a theory. It has no explanatory usefulness by itself, and there is no causal relationship, or isomorphic correspondence, between model and theory. A model aids understanding by explicitly delineating structural relations in a theory, although Meehan [22] emphasizes that extensive reliance upon a model leads often to an overemphasis on rigor and precision and a loss of understanding. A model is not used in an exploratory sense to discover underlying generalizations nor is it used in the theoretical sense to explain general statements. It is used to depict the structural bones of a system, and that is how it is used in GEMLUP.

The form of the model used to depict induced development in GEMLUP is path analysis, a causal relationship. This is in keeping with factor theory, which is also a causal approach to explanation.

## 2. Causal Analysis

The first topic to be addressed in this section is causality and causal thinking. The need for, and limitations of, causal relations in modeling are also addressed. The second topic is the connection between causal thinking and modeling and path analysis.

### a. Causality

Causality, or the cause-effect relationship, is an important aspect of science. Classical experimentation involves isolating one independent variable from all others in a relationship, varying it, and recording the resulting change (if any) on the dependent variable of interest. Change in the dependent variable is then assumed to be caused by change in the independent variable, as all other variables were held constant. While systems analysis has altered this one-to-one isolated scientific approach, causality is still the dominant philosophical principle in scientific research[26].*

---

*There are many kinds of theory. For instance, Meehan [22] discusses the following kinds of theory: deductive (pure deductive and probabilistic deductive), concatenated, instrumental, factor, and quasi-theory.

"The notion of causality applies when the occurrence of one event is reason enough to expect the production of another" [19]. The word itself is used in three ways: (1) as a category to denote a particular bond relationship between a cause and an effect, (2) as a principle, or general law, relating objects, and (3) as a doctrine which purports that causality is the only form of determinism operating in reality [26]. We will only be concerned in GEMLUP projects with the first use of the word.

Philosophers of science interested in causality do not believe that the phenomenon holds 100% of the time in all circumstances. They are much more modest in their claims. For instance, Mario Bunge in Causality states that causal chains work best as rough approximations for relatively short periods of time. They are valid in some contexts because causality is a rough reflection of reality and is methodologically unavoidable: there is just no alternative available that can be used for explanation and prediction [26].

There is a place for chance variations, random impacts, and side reactions in causality [26]. The idea is that even if these effects do not cancel out each other, their net effect is statistically tractable. There is also a place for multiple causes, multiple effects, and mutual causality (feedback). Their impact on a causal relation can be determined statistically or, in the case of feedback, by disaggregation into asymmetrical causal stages. What cannot be handled very well in causality is systematic interaction among causes or effects. The reason for this is that interaction introduces non-additivity into the system, which cannot be handled in a causal analysis. (Which is based upon the premise that the joint effect of many forces, or causes, is equal to the vector sum of separate forces.) If interaction is suspected to be a property of the system under study, it may be possible to approximate this relation by a relatively simple transformation (e.g., a multiplicative model, or a logarithmic transformation) so that causal modeling can be used on the system [27]. If a transformation is not theoretically reasonable, another form of analysis must be used for the interactive system. Not everything can be causally modeled, in other words.

---

*A formal definition of causation appears in David Heise, Causal Analysis [15]. An event C, causes another event, E, if and only if

(a) an operator exists which generates E, which responds to C, and which is organized so that the connection between C and E can be analyzed into a sequence of compatible components with overlapping event fields;
(b) occurrences of event C are coordinated with the presence of such an operator: such an operator exists within the field of C;
(c) when conditions (a) and (b) are met, when the operator is isolated from the fields of events other than C, and neither C nor E is present to begin with, then occurrences of C invariable start before the beginning of an occurrence of E;
(d) when conditions (a) and (b) are met, C implies E; that is, during some time interval occurrences of C are always accompanied by occurrences of E, though E may be present without C or both events may be absent.

Additional drawbacks of causal analysis are (1) the treatment of (causal) relations in isolation from the environment as separate events, and (2) the downplaying of time in the system [26]. The first drawback is common to most systematic attempts to model reality. The whole world cannot be mathematically modeled, so simplications have to be made throughout the analysis. This results in an artificial boundary being placed around a system, with the emphasis being placed on what goes on inside the boundary rather than on the interfaces between system and environment [28]. Not only that, but causal analysis focuses on isolated events within the system, again oversimplifying reality. (Bunge calls this ontologically objectionable.) It is analytic and not synthetic, in other words.

Time is usually not explicitly handled in most causal analyses, other than the logical time relation inherent in the definition of causality: "an event is not caused by other events that occur after it" [15]. (This is the so-called temporal ordering criterion of causality.) This situation is due more to lack of knowledge and theory regarding time relations in social systems than an inability of causal analysis to incorporate time in its methods. In other words, causal analyses can use "higher level" techniques to get time into the equations (finite differencing) so that meaningful conclusions can be drawn concerning the results. We felt that GEMLUP models had to be static and not dynamic for these reasons and one other--money. Obtaining longitudinal data is a very expensive process and is well beyond the Land Use Planning Office's budget. Hence, the use of static cross-sectional path anlaysis to test the causal model [25].

### b. Path Analysis

Path analysis is a systems approach to testing causal relations in an operationalized theory (a model). It is one of a number of partial-correlation analysis techniques used for modeling a system of interrelated variables that ultimately depend upon outside (exogenous) variables to "drive" them [29]. The technique has found its greatest application in sociology and econometrics, although it was first developed by a biologist, Sewall Wright (in 1934). Lucid explanations of path analysis are found in Van de Geer [29], Blalock [28], Heise [15], and Benesh [17]. The last citation is Volume I of the GEMLUP reports. Chapter II of the Volume, written by Dr. Ralph D'Agostino of Boston University, contains a good summary of the technique taken from the references just noted; the Chapter is reproduced, in updated form, in Appendix A of this report.

The only attribute of path analysis that will be discussed here is its ability to "tease out" causal relations. This ability is the reason why we use it in GEMLUP projects--and the reason why we can state that the land uses identified as being secondary development induced by constructing and operating a major land use project are just that. It is the way in which "disturbance" is controlled in the land use development system so that specific cause and effect relations can be identified [15]. And this ability to tease out causal relations via partial-correlation provides us with the answer to the question posited earlier: "How do you know that the land uses you are predicting are really secondary development induced by the major project?" The answer is, finally, partial-correlation analysis as used in a systems context via path analysis allows us to infer causality (with a known probability of error), or in our context, allows us to infer induced development. This sentence is the short answer to the short question. But we are not through yet, as more should be said concerning partial-correlation analysis. (See also the material on partial and semi-partial correlation presented earlier under "Prediction, Explanation, Theory, and Model.")

Heise presents the case most clearly in a discussion of separating out the individual effects of multiple causes. The mish-mash of multiple causation is really what the question of interest reproduced above is all about. The question can be rephrased as: "Given that some development has occurred in the area of interest after the major project was built, how do you know it was induced, or caused, by the major project and not something else?" Obviously, multiple causation is the issue here, and we turn to Heise for elucidation [15].

> Multiple causation complicates causal analysis because it creates a situation in which the value of an effect is not determined solely by the cause of focal interest but by other causes as well. When trying to study the relation between a specific cause and effect, all the other causes act as disturbing factors that confound and frustrate analysis. To study the relationship of interest the disturbing factors somehow must by controlled. Three different strategies are available.
>
> One approach, associated mainly with classical experimentation in the physical sciences, involves isolation of the relationship of interest. Any disturbance depends on a causal relationship, so disturbances occur because of the existence of some operator coordinated with some source event. Therefore, if the operator for a disturbance can be disconnected, made inoperative, or insulated from its stimulus events, this cause-effect relationship no longer exerts its influence. Moreover, if the causal relations for all disturbing factors are disrupted, the relation between the remaining cause and effect can be analyzed without worrying about disturbances. . .Employment of this strategy is not always so simple as it may seem. It takes considerable skill and knowledge to disrupt every conceivable operator or to control all unwanted events. Indeed the strategy may be applicable only in problems in which a great deal of scientific knowledge already exists to guide experimental design.
>
> A second method of dealing with the problem of disturbances is to establish a kind of passive control over the disturbing factors by observing them in detail to determine when one or another is creating an unwanted effect so that it can be discounted or otherwise adjusted. In its simplest form this would mean restricting analysis to periods in which disturbing factors are not creating disturbances. . .
>
> Third, the problem of disturbing factors can be approached from a statistical perspective. It is accepted that any single observation is hopelessly confounded by disturbing factors, but if enough instances are observed in which the presumed source operates at certain levels it should be possible to determine whether the source has the proposed effect on the average. In averaging over many cases, the effects of the disturbing factors hopefully will cancel one another. . .

> A key requirement here is that the disturbing factors must be unrelated to the causal event of interest. If this were not true--if one disturbing event always tended to occur with the causal event of interest--then the effect of that disturbance would not disappear with averaging. Furthermore, it is always necessary to obtain observations with the presumed cause both present and absent, or set at various levels, in order to ascertain whether the presumed source makes any difference.
>
> The statistical averaging idea has been elaborated in two different directions. The most rigorous procedure (but frequently impractical in social situations) is the statistical experiment in which a lack of coordination between disturbing factors and the presumed cause of interest is guaranteed by randomly assigning cases for observation onto two different sets and then providing the hypothesized causal event for just one of them. Then, because the presumed cause is uncoordinated with any of the possible disturbing factors, its effect can be assessed by finding the average magnitude when the presumed cause is absent.
>
> In the second approach observations of disturbing factors are used to define groups of measurements in which each disturbance has a constant value. Then the value of the effect variable in each group is reset arithmetically to the magnitude it would have if disturbances had not been present, whereupon the relation of interest can be examined to see whether there is correspondence between the hypothesized cause and the adjusted effect indicating causality. In actual applications the adjustment approach really amounts to studying several causal relationships simultaneously. The procedure for adjusting for the impact of a second cause while studying a first requires making an adjustment for cause one in order to determine how much to adjust for cause two. At first, this sounds hopelessly circular and complex. In fact, procedures exist for making this separation, given numerous observations on all relevant variables. . . .

Heise then references two chapters in his book for elaboration of the procedures. The first deals with the logic behind causal thinking and the second describes methods to identify and estimate causal relations in path analysis. The latter boils down to partial and semi-partial correlation analysis. This is how statistical control is used to infer causality. This is why we do path analytic modeling in GEMLUP.*

---

*If the logic is not clear here or if the reader is still not convinced, please read Heise [15] and Kerlinger and Predhazur [6]. Blalock [27] and Van de Geer [29] also discuss the issue, but not as directly.

## C. AERIAL PHOTOGRAPHIC INTERPRETATION

As part of the data collection for this study, it was necessary to inventory land use by category, in the case study areas of analysis. Although some regions were found to maintain detailed land use maps and inventories, these were the exception. Therefore, this information was generally extracted from aerial photographs of the case study areas. The experience from the GEMLUP-I project [17] indicated that this element of the data collection had the greatest potential for measurement errors. Thus, the objective of this task was to determine the "typical" errors associated with estimating amounts of land use, by category, from aerial photographs. The approach involved a comprehensive review and compilation of information in the technical literature.

### 1. Literature Review

A total of 33 sources were reviewed. Telephone conversations with personnel in the U.S. Geological Survey (USGS) Geography Applications Program supplemented these sources. Of these sources, twelve did not indicate the level of accuracy that would result from using their recommended procedures for interpreting aerial photographs. The remaining sources offered no consistency with regard to their method of land use interpretation, scale of the photograph used, or photographic equipment. A summary of these sources is indicated in Table C-1.

As noted in Table C-1, the range in photographic scale reported in the literature sources reviewed ranged from 1:2,400 to 1:250,000.* In addition, the methods of interpretation varied, including ground checks by windshield surveys of remote sensing data [31], center point checks of 1-km grid cells [32], field comparisons along linear traverses [32], the interpretation of items which could only be checked through ground examinations [33], and the direct examination of color infrared photographs [34]. Further, the sources varied with regard to the nature of the area to be interpreted. For example, Lindgren [34] concentrated on residential areas; Collins and El-Beik [35] dealt solely with a portion of the City of Leeds, England; Branch [33] examined a minor portion of Baldwin Park, California; and Fitzpatrick [32] covered the Central Atlantic Coastal Region stretching from the southern border of Virginia to the southern quarter of New Jersey. In addition, the literature reveals that a considerable variety of land use classification schemes were employed. Collins and El-Beik [35] classified urban land use according to eight major groups; Fitzpatrick [32] used a revised USGS Level I and II classification system; and Mausel et al. [31] adopted a system of earth features that could be delineated by machine assisted processing of LANDSAT-1 multispectral data. Finally, Simpson [36] based his

---

* The ratio expresses the size relationship between features on a map and the same features on the earth's surface. Thus, 1:2,400 states that 1 inch on the map represents 2,400 inches (or 200 feet) on the earth's surface.

**Preceding page blank**

TABLE C-1

SUMMARY OF AERIAL PHOTOGRAPHIC ACCURACY DETERMINATION

| Authors | Overall Accuracy | Most Accurate Category | Least Accurate Category | Comments |
|---|---|---|---|---|
| Mausel, Leivo & Lewellen [31] | 85-90% | Water (98%) |  | Remote sensing data checked by windshield survey |
| Harden [37] | 90% |  |  | Rule of thumb based on cumulative experience |
| Collins and El-Beik [35] | 92% | Water bodies (100%)<br>Transportation (100%)<br>Open improved lands (99%)<br>Open unimproved lands (96%)<br>Residential (92%) | Complex urban land use (36.4%)<br>Commercial (84.5%)<br>Industrial (86.5%) | Stereoscope examination of 1:10,560 scale aerial photos with ground check |
| Fitzpatrick [41] |  | Residential (94%)<br>Cropland & Pasture (93%)<br>Mixed Forest (91%)<br>Exergreen forest (90%) | Deciduous forest (78%) | USGS study of Greater Atlanta Region (in draft form) |
| Fitzpatrick (CARETS) [32] | 84.9% | Bays & Estuaries (89.9%)<br>Cropland & Pasture (83.6%) | Residential (74.5%) | Point check of 1:24,000 scale aerial photos |
| Fitzpatrick (CARETS) [32] | 91% |  | Light Crown Cover Forest (40%) | Linear Traverse Technique at 1:100,000 |
| Fitzpatrick (CARETS) [32] | 69.5% |  | Urban (57%) | ERTS maps at 1:250,000 |
| Fitzpatrick (CARETS) [32] | 76.5% |  |  | High-altitude photos 1:250,000 |
| Lindgren [34] | 99.5% |  | Dwelling units per residential structure (59%) | Color infrared photos of residential areas 1:20,000 |
| Simpson [36] | 81.3% |  |  | Sample test of interpreters |
| Branch [33] | 96.5% |  |  | Included items that could only be ground checked 1:2,400 |
| Avery [39] | 90% |  |  | Rule of thumb |

determination of overall accuracy on a test involving "the 25 most competent radar interpreters available nationally..."while Harden [37], Fitzpatrick [38], and Avery [39] based their conclusions on a "rule of thumb" which they felt was an acceptable level of overall accuracy to be attained by professional interpreters of aerial photographs.

Three sources were especially relevant. Collins and El-Beik [35] indicated that they used 1:10,560 scale aerial photographs which were of poor quality. Regardless, they achieved an overall level of accuracy of 92%. The percentage of correctly identified land uses ranged from 36.4% for complex urban land use (i.e., high density mixed uses) to near 100% for bodies of water and transportation. More revealing, however, was the percentage over or under estimation by land use category. Commercial areas were over-estimated by 24%, transportation was over-estimated 22%; and complex urban land use was under-estimated by 39.4%.

By contrast, a limited check of aerial photo data used in the GEMLUP-I project [17] revealed that commercial land use was under-estimated by 12%, manufacturing over-estimated by 10%, industrial over-estimated by 71% and office land use under-estimated by 40%.

In the CARETS project, Fitzpatrick [32] reported the accuracy of determining land use by categories according to three scales of aerial photos (see Table C-2). Figure C-1 depicts Fitzpatrick's findings in the form of a bar chart. The obvious was concluded, i.e., the more detailed the aerial photograph, the greater the degree of accuracy per land use category. Two exceptions, however, were noted. First, urban and built-up land was most accurately mapped at a scale of 1:100,000. Second, nonforested wetlands were mapped most accurately at a scale of 1:24,000 and 1:250,000.

## 2. Interpretation Factors

On the basis of the literature and professional experience, the following factors are of significance in interpreting aerial photographs:

1. Skill of the interpreter (Ability to decipher land use from aerials on the basis of tone, texture, color, shadows, slope, pattern, size, shape, and ralationship to other features).
2. The quality of the photograph.
3. The scale of the photograph.
4. The season when the photograph was taken.
5. The equipment used for taking and interpreting the photograph.
6. The extent to which the form of the photograph is reflected in its function.
7. The nature of the area of analysis.

## TABLE C-2

## COMPARISON OF AREA FOR VARIOUS CATEGORIES OF LAND USE MAPPED AT THREE SCALES FROM HIGH-ALTITUDE PHOTOGRAPHY (IN HECTARES)

Source: [32]

|  | L.U. CODE | 1:24,000 | 1:100,000 | 1:250,000 |
|---|---|---|---|---|
| URBAN AND BUILT-UP |  |  |  |  |
| Residential | 11 | 4,069 | 3,441 | 4,950 |
| Commercial and services | 12 | 340 | 464 | 300 |
| Industrial | 13 | 38 | 48 |  |
| Extractive | 14 | 33 | 112 |  |
| Transportation, etc. | 15 | 259 | 132 | 225 |
| Institutional | 16 | 961 | 1,289 | 875 |
| Strip and clustered | 17 | 60 | 48 |  |
| Mixed | 18 |  | 16 |  |
| Open and other | 19 | 409 | 400 | 325 |
| Subtotal | 1 | 6,169 | 5,950 | 6,675 |
| AGRICULTURAL |  |  |  |  |
| Cropland and pasture | 21 | 21,544 | 23,156 | 23,875 |
| Orchards, etc. | 22 | 10 | 404 |  |
| Other | 24 | 95 | 92 |  |
| Subtotal | 2 | 21,649 | 23,652 | 23,875 |
| FOREST LAND |  |  |  |  |
| Heavy crown cover | 41 | 33,740 | 31,550 | 30,950 |
| Light crown cover | 42 | 1,217 | 1,906 | 1,900 |
| Subtotal | 4 | 34,957 | 33,456 | 32,850 |
| WATER |  |  |  |  |
| Streams and waterways | 51 | 334 | 404 | 75 |
| Lakes | 52 |  | 108 |  |
| Reservoirs | 53 | 224 | 92 | 125 |
| Bays and estuaries | 54 | 9,316 | 9,150 | 8,850 |
| Subtotal | 5 | 9,874 | 9,754 | 9,050 |
| NONFORESTED WETLAND |  |  |  |  |
| Vegetated | 61 | 3,273 | 3,088 | 3,500 |
| Subtotal | 6 | 3,273 | 3,088 | 3,500 |
| BARREN LAND |  |  |  |  |
| Sand other than beaches | 72 | 6 |  |  |
| Beaches | 74 | 10 |  |  |
| Other | 75 | 62 | 100 | 50 |
| Subtotal | 7 | 78 | 100 | 50 |
| TOTAL |  | 76,000 | 76,000 | 76,000 |

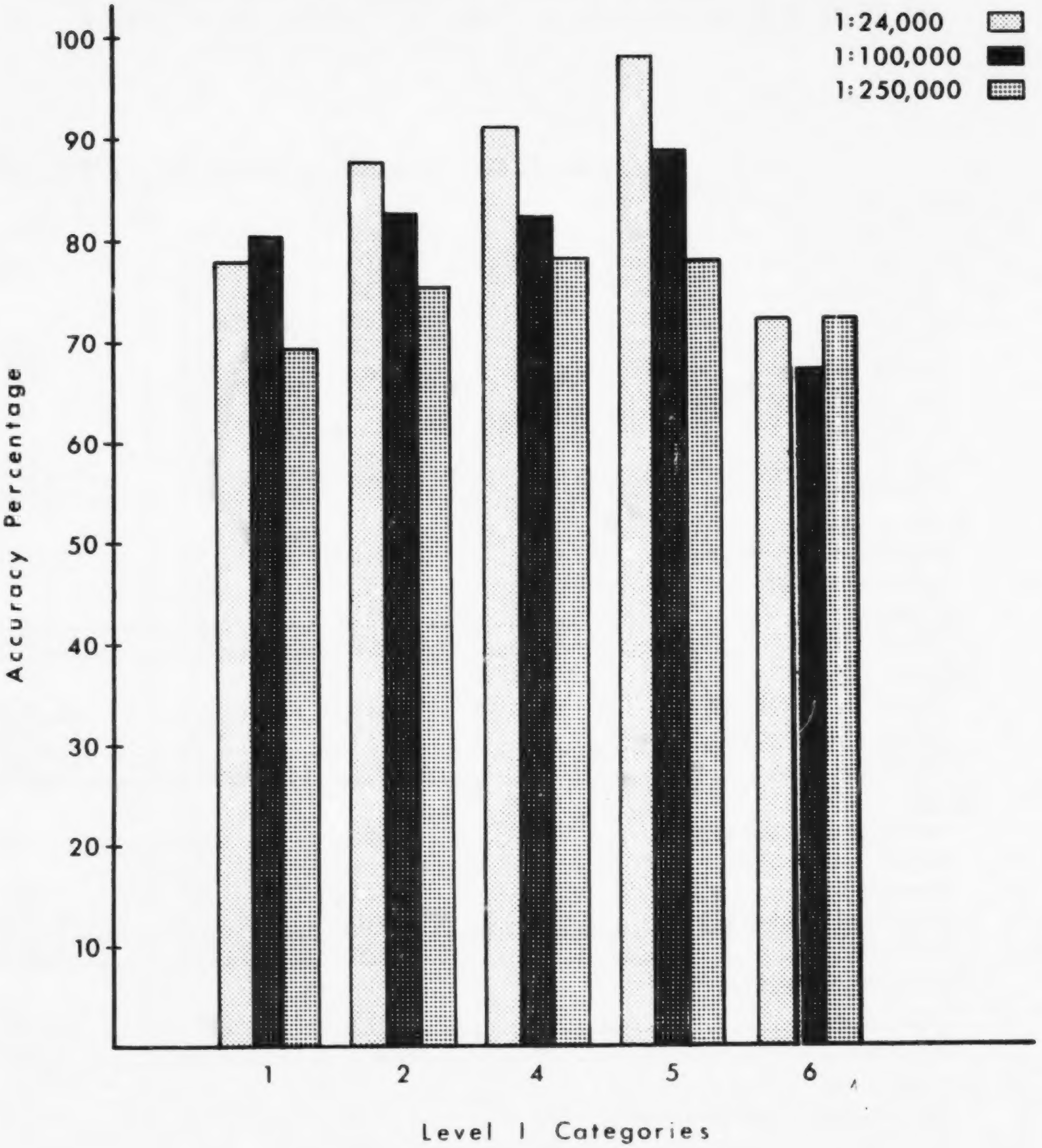


FIGURE C-1

A COMPARISON OF THE ACCURACY OF LEVEL I LAND USE INTERPRETATIONS AT THREE SCALES DERIVED FROM AIRCRAFT DATA FOR EACH LAND USE CATEGORY. PERCENTAGES DERIVED FROM FIELD CHECK.

Source: [32]

8. The land use classification scheme employed (Categories should be mutually exclusive).
9. The resolution desired.
10. The availability of supplemental materials.
11. Repeatability by other interpreters.

Of these factors, the most important are (1) the skill of the interpreter, (2) the scale of the photograph, (3) the land use classification system employed. With regard to the first factor, Simpson [36] reported--as indicated above--that 25 competent interpreters averaged 81.3% accuracy. The U.S. Department of Agriculture has reported [40] two instances where skilled soil scientists with 8 and 12 years experience achieved 82.5% and 95% accuracy, respectively. Second, the scale of the photograph is generally significant vis-a-vis the level of accuracy. As previously indicated, Fitzpatrick [32] reported that the accuracy of land use interpretation increased as aerial photo scale increased. By way of another example, Lindgren [34] concentrated on residential areas using photos with a scale of 1:20,000 and attained an accuracy level of 99.5%. Finally, the land use classification system employed should utilize categories that are mutually exclusive so as to minimize subjective judgments. The difficulty here is that such classification schemes will vary according to the area under examination, the topics of particular interest to the interpreter and limitations of the equipment used.

### 3. Conclusions: Level of Accuracy

Based on the review of aerial photographic literature, the average overall level of accuracy--regardless of the numerous variables involved such as scale, equipment, etc.--was 87%. This is slightly lower than the 90% general "rule of thumb" mentioned by Harden [37], Fitzpatrick [38], and Avery [39]. This "rule of thumb", however, should be taken as an ideal to be achieved, i.e., accuracy will typically not exceed this level.

The land use categories that typically attained the highest level of accuracy were bodies of water (94.5%) [35]; open improved and unimproved land (97.5%) [35]; and mixed forest (91%) [41]. The land use categories for which the least accuracy was achieved included complex urban use (36.4%) [35]; urban land use taken from ERTS photographs at a scale of 1:250,000 (57%) [32]; dwelling units per residential structure (59%) [34]; and light crown forest cover (40%) [32]. Urban land use categories achieved the following accuracies: residential (59%-94%) [32,34,35,41], commercial (84.5%) [35], Industrial (86.5%) [35] and general urban land use (36.4% - 57%) [32,35]. Due to the wide variability in accuracy figures, by category, the lack of sufficient literature on interpretation of urban land use (as opposed to vegetation covers), and the numerous intervening factors, the results of the literature review were inconclusive and "typical" errors associated with each land use category could not be determined. Instead, the average overall level of accuracy--87%--is recommended. This figure is used in the coefficient stability analysis of the causal model (see Section IV).

## D. BIBLIOGRAPHY

This section summarizes all relevant citations uncovered in the literature search task of this study. The literature search was aimed at identifying publications in the following four technical areas:

- Causal Analysis and Predictive Modeling of Induced Development Associated with Wastewater Facilities
- Simplified VMT Predictive Modeling Procedures
- Land Use Identification and Categorization from Aerial Photographs
- Potential Case Study Wastewater Projects

### 1. Causal Analyses and Predictive Modeling of Induced Development Associated with Wastewater Facilities.

Abt Associates, Manual for Evaluating Secondary Impacts of Wastewater Treatment Facilities (Draft), Cambridge, MA, 1976.

Bascom, S.E., Cooper, K.G., Howell, M.D., Makcrides, A.C., and Rabe, F.T., Secondary Impacts of Transportation and Wastewater Investments: Research Results (EPA-600/5-75-013) and Review and Bibliography (EPA-600/5-75-002), Washington, DC, 1975.

Benesh, F., Guldberg, P., and D'Agostino, R., Growth Effects of Major Land Use Projects: Volume I - Specification and Causal Analysis of Model, EPA Publication No. EPA-450/3-76-012a, May 1976.

Benesh, F., Guldberg, P., and D'Agostino, R., Growth Effects of Major Land Use Projects: Volume III - Summary, EPA Publication No. EPA-450/3-76-012c, Research Triangle Park, NC; September 1976.

Binkley, C., et al., Interceptor Sewers and Urban Sprawl, Lexington Books, Lexington, MA, 1975.

Blalock, H., Theory Construction, Englewood Cliffs, NJ: Prentice-Hall, Englewood Cliffs, NJ, 1969.

Blatnik, J.A., "History of Federal Pollution Control Legislation", Industrial Pollution Control Handbook, McGraw-Hill, New York, NY, 1971.

Booz-Allen & Hamilton, Inc., Methodologies for the Analysis of Secondary Air Quality Impacts of Wastewater Projects Located in AQMAs, Environmental Impact Office, EPA Region II, New York, NY, March 1976.

**Preceding page blank**

Boudon, R., "A Method of Linear Causal Analysis: Dependence Analysis", American Sociological Review 33:365-373, 1965.

Bureau of Regional Planning, Secondary Impact of Regional Sewerage Systems, State of New Jersey Department of Community Affairs, Trenton, NJ, 1975.

Computer Sciences Corporation, Infonet Times Series Processor, Information Network Division, Los Angeles, CA, December 1974.

Council on Environmental Quality, Guidelines for Preparation of Environmental Impact Statements, 38 Federal Register (August 1, 1973).

Council on Environmental Quality, The Fifth Annual Report of the Council on Environmental Quality, Washington, DC, 1974.

Environmental Impact Center, Secondary Effects of Public Investments in Highways and Sewers, Washington, D.C., Council on Environmental Quality, 1974.

Environmental Impact Center, Secondary Impacts of Infrastructure Investments in the Denver Region, Washington, D.C., Council on Environmental Quality, 1974.

Fensterstock, J.C. and Speaker, D.M., Use of Environmental Analyses on Wastewater Facilities by Local Government, Washington, D.C., U.S. Environmental Protection Agency, 1974, (EPA-600/5-74-015).

Goldberger, A.S., Econometric Theory, Wiley, New York, NY, 1964.

Grava, S., Urban Planning Aspects of Water Pollution Control, Columbia University Press, New York, NY, 1969.

Heise, D., Causal Analysis, John Wiley & Sons, New York, NY, 1976.

Heise, D.R., "Problems in Path Analysis and Causal Inference", Sociological Methodology, edited by E.F. Borgatta, Josey-Bass, San Francisco, CA, 1969 (pp. 38-72).

Hill, D., "A growth allocation model for the Boston region", Journal of the American Institute of Planners, 31(2), 1965.

Hudson, J.F., Demand for Municipal Services: Measuring the Effect of Service Quality, Ph.D. thesis, MIT Department of Civil Engineering, Cambridge, MA, June 1975.

Johnston, J., Econometric Methods, McGraw-Hill, New York, NY, 1963.

Kenney, Kenneth B., Downing, Donald A., and Hayes, Gary G., Urban Water Policy as an Input in Urban Growth Policy, Water Resources Research Center, Tennessee University, Knoxville, TN, 1972 .

Kerlinger, F.N., and Pedhazer, E.J., Multiple Regression in Behavioral Research, Holt, Rinehart and Winston, New York, NY, 1973.

Land, K.C., "Principles of Path Analysis", Sociological Methodology, Edited by E.F. Borgatta, Jossey-Bass, San Francisco, CA, 1969.

Milgram, G., The City Expands - A Study of the Conversion of Land from Rural to Urban Use in Philadelphia, 1945-62, University of Pennsylvania, Philadelphia, PA, March 1967.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., and Bent, D.H., Statistical Package for the Social Sciences (second edition), McGraw-Hill, New York, NY, 1975.

Pepper, James E. and Jorgensen, Robert E., Influences of Wastewater Management on Land Use: Tahoe Basin, Washington, DC, U.S. Environmental Protection Agency, 1974 (EPA-600/5-74-019; NTIS #PB-240-247/7ST).

Phillips, M.B., "Developments in Water Quality and Land Use Planning: Problems in the Application of the Federal Water Pollution Control Act Amendments of 1972". Urban Law Annual, 10:43, 1975.

Promise, J., "Dallas and D.C. - An Insider's Comparison of '208' Approaches to Growth Policy", 49th Annual Conference of the Water Pollution Control Federation, Minneapolis, MN, October 1976.

Promise, J. and Leiserson, M., Water Resources Management for Metropolitan Washington: Analysis of the Joint Interactions of Water and Sewer Service, Public Policy, and Land Development Patterns in an Expanding Metropolitan Area (including Appendices), Washington, DC, Metropolitan Washington Council of Governments, 1973.

Rabe, F.T. and Hudson, J.F., "Highway and Sewer Impacts on Urban Development:, J. Urban Planning and Development - ASCE, 101:217, November 1975.

Real Estate Research Corporation, The Costs of Sprawl - Environmental and Economic Costs of Alternative Residential Development Patterns at the Urban Fringe, prepared for the Council on Environmental Quality, Washington, DC, April 1974.

Reid, George W. and Alquire, Robert T., Systems Approach to Metropolitan and Regional Area Water and Sewer Planning, Norman, OK, Bureau of Water Resources Research, Oklahoma University, 1973 (NTIS #PB-222-262/8).

Rivkin/Carson, Inc., Population Growth in Communities in Relation to Water Resources Policy, prepared for the National Water Commission, Arlington, VA, October 1971.

Snedecor, G.W., and Cochran, W.G., Statistical Methods (6th Edition), Iowa State College Press, Ames, IA, 1967.

Stansbury, J., "Suburban Growth - A Case Study", Population Bulletin, 28:5, February 1972.

Stokes, E.B., "The Evaluation of an EMPIRIC Model for an Urban Area", Environment & Planning A, pp. 703-715, 1973.

Tabors, R., Shapiro, M., and Rogers, P., Land Use and the Pipe: Planning for Sewerage, Heath, Lexington, MA, 1976.

Thomas, R.D., Water Problems in the Context of County Government Decision-Making, Water Resources Research Center Publication No. 32, Florida Atlantic University, Boca Raton, FL, October 1975.

Train, Russell E., "Memo on the Consideration of Secondary Environmental Effects of Construction Grants" (Program Guidance Memorandum #50), Washington, DC, U.S. Environmental Protection Agency, June 7, 1975.

Train, R.E., "The EPA Programs and Land Use Planning", Columbia Journal of Environmental Law, 2:255, Columbia University, New York, NY Spring 1976.

Turner, M.E., and Stevens, C.D., "The Regression Analysis of Causal Paths", Biometrics, 15:236-258, 1959.

Tukey, J.W., "Causation, Regression and Path Analysis", Statistics and Mathematics in Biology, Edited by O. Kempthorne, T.A. Bancroft, J.W. Gowen and J.L. Lush, Iowa State College Press, Ames, IA, 1964, pp. 35-66.

Urban Systems Research and Engineering, Inc., Interceptor Sewers and Suburban Sprawl: The Impact and Construction Grants on Residential Land Use, Volume I: Analysis, and Volume II: Case Studies, prepared for the Council on Environmental Quality, Washington, DC, September 1974.

Urban Systems Research & Engineering, Inc., The Distribution of Water Pollution Control Costs, prepared for the National Commission on Water Quality, Washington, DC, 1976.

Urban Systems Research & Engineering, Inc., The Growth Shapers - The Land Use Impacts of Infrastructure Investments, prepared for the Council on Environmental Quality, Washington, DC, May 1976.

U.S. Congress, Federal Water Pollution Control Act (33 USC 1251 et. seq.).

U.S. Environmental Protection Agency, "Annotated Bibliography for Water Quality Management" (3rd Edition), Water Planning Division, October 1976.

U.S. Environmental Protection Agency, "Annotated Bibliography - Publications and Active Projects", Comprehensive Planning and Land Use Staff, Washington, DC, May 1976.

U.S. Environmental Protection Agency, Guidelines for State and Areawide Water Quality Management Program Development (Draft), Washington, DC, Water Programs Division, February 1976.

U.S. Environmental Protection Agency, "Land Use and Water Quality", Proceedings of the EPA Pollution Control Technology Assessment Conference, Columbus, OH, May 1974.

U.S. Environmental Protection Agency, Mitigating Secondary Impacts from the Wastewater Facilities Program, Office of Land Use Coordination, Washington, DC, 1977.

U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Guidelines for Air Quality Maintenance Planning and Analysis, Research Triangle Park, NC, 1974, Volume 3: Control Strategies (EPA-450/4-74-003), Volume 4: Land Use and Transportation Considerations (EPA-450/4-74-004), Volume 7: Projecting County Emissions (EPA-450/4-74-008), Volume 9: Evaluating Indirect Sources (EPA-450/4-75-001), Volume 13: Allocating Projected Emissions to Sub-County Areas (EPA-450/4-74-014).

U.S. Environmental Protection Agency, Policies and Procedures for Continuing Planning Process (40 CFR 130), reprinted in 40 Federal Register 55334; November 28, 1975; also reprinted in: Bureau of National Affairs, Environmental Reporter; Federal Regulations (131:2673).

U.S. Environmental Protection Agency, Regulations on Preparation of Water Quality Management Plans (40 CFR 131), reprinted in 40 Federal Register 55343; November 28, 1975; also reprinted in: Bureau of National Affairs, Environmental Reporter; Federal Regulations (131:2681).

Van de Geer, J.P., Introduction to Multivariate Analysis for the Social Sciences, W.H. Freeman, San Francisco, CA, 1971.

Wonnacott, R.J., and Wonnacott, T.H., Econometrics, John Wiley & Sons, Inc., New York, NY, 1970.

Wright, S., "Path Coefficients and Path Regressions Alternative or Complementary Concepts", Biometrics, 16:189, 1960.

Wright, S., "The Interpretation of Multivariate Systems", Statistics and Mathematics in Biology, Edited by O. Kempthorne, T. Bancroft, J. Cowen, and J. Lusn, Iowa State College Press, Ames, IA, 1954 (Vol. 11-33).

Wright, S., "The Method of Path Coefficients", Annals of Mathematical Statistics, 5:161-215, 1934.

Wright, S., "The Treatment of Reciprocal Interaction, with or without Lag, in Path Analysis", Biometrics, 16:423, 1960.

## 2. Simplified VMT Predictive Modeling Procedures

Ashford, N. and Holloway, F.M., The Permanence of Trip Generation Equations, Florida State University, 1971, NTIS PB-204-433.

Australian Road Research Board, Proceedings of the Sixth ARRB Conference, 6 (2): 5-22, Melbourne, Australia, 1972.

Brand, D., "Theory and Method in Land Use and Travel Forecasting", Highway Research Record #422, pp. 10-20, 1973.

Burke, R.H., Atkins, A.S. and Coote, G.M., Procedures for Forecasting Vehicle Miles of Travel in National Road Planning, Australian Bureau of Roads, Melbourne, Australia, 1974.

Cesario, F.J., "Combined Trip Generation and Distribution Model," Transportation Science, 9 (3):211-223, August 1975.

Chang, H.K., and Smith, C.L., Trips Ends Generation Research Counts, California State Department of Transportation, July 1973, NTIS PB-227 134/4.

Charles River Associates, A Disaggregated Behavioral Model of Urban Travel Demand, Cambridge, MA, March 1972 NTIS PB-210 515.

Chatterjee, A. and Cribbins, P.D., "Forecasting Travel on Regional Highway Network", ASCE Transportation Engineering Journal, pp 209-224, May 1972.

DeLeuw, Cather, Canada Ltd., A New Procedure for Urban Transportation Planning, Department of Highways, Ontario, Canada, September 1969.

Devlin, J., Hearne, R., and McGuinness, P., Vehicle Miles Of Travel For 1976, The National Institute for Physical Planning and Construction Research, Dublin, Ireland, 1977.

Federal Highway Administration, Guide For Forecasting Traffic On The Interstate System, U.S. Department of Transportation, Highway Planning Program Manual #134, Washington, DC, March 1973.

Federal Highway Administration, Guidelines For Trip Generation Analysis, Washington, DC, April 1973, NTIS PB-244 925/4ST.

Federal Highway Administration, Nationwide Personal Transportation Study, Report #8, "Home to Work Trips and Travel", Washington, DC, August 1973, NTIS PB-242 892/8ST.

Federal Highway Administration, Nationwide Personal Transportation Study, Report #10, "Purposes of Automobile Trips and Travel", Washington, DC, May 1974, NTIS PB-242 894/4ST.

Federal Highway Administration, Statewide Travel Demand Forecasting, U.S. Department of Transportation, Highway Planning Program Manual #147, Washington, DC, November 1973.

Federal Highway Administration, Travel Simulation For Small Cities (Draft), U.S. Department of Transportation, Washington, DC, 1973.

Fogarty, W.J., "Trip Production Forecasting Models For Urban Areas", ASCE Transportation Engineering Journal, pp 831-845, November 1976.

Gustafson, R.L., "Empirical Study of Factors Influencing Trip Attraction and Trip Generation", High Speed Ground Transportation Journal, 7 (3):307-321, 1973.

Highway Research Board, Urban Travel Demand Forecasting, Transportation Research Board Special Report #143, 1973.

Jeffries, W.R. and Carter, E.L., Simplified Techniques For Developing Transportation Plans - Trip Generation in Small Urban Areas, West Virginia University, 1966, NTIS PB-183 217.

Kassoff, H., and Gendell, D.S., "An Approach to Multiregional Urban Transportation Policy Planning", Highway Research Record #348, Highway Research Board, Washington, DC, 1971.

Keefer, L.E. and Witheford, D.K., Urban Travel Patterns for Hospitals, Unversities, Office Buildings, and Capitols, National Cooperative Highway Research Program Report #62, 1969.

Landis, R., The Effect of Automotive Fuel Conservation Measures on Air Pollution, EPA Publication No. EPA-600/5-76-006, Washington, DC, September 1976.

Mann, W.W., TRIMS - A Procedure For Quick Response Transportation Planning, Metropolitan Washington Council of Governments, Washington, DC, January 1975.

Maricopa Association of Governments, Trip Generation By Land Use, Washington, DC, April 1975, NTIS PB-244 549/2ST.

Meyerowitz, W., Trip Generation: Regression Analysis, Procedural Manual, Michigan Department of State Highways, 1970, NTIS PB-198 506.

Morris, R.J., A Comparative Analysis of Trip Distribution and Traffic Assignment Models for Transportation Planning in Developing Areas, Stanford University, 1973, NTIS PB-232 325/1.

NCHRP, Factors, Trends, and Guidelines Related to Trip Length, National Academy of Engineers Report #89, Washington, DC, 1970.

Schneider, M., Transportation and Land Development: Unified Theory and Prototype Model, Creighton Hamburg Inc., 1969, NTIS PB-184 941.

Soliman, A. and Sharma, S.C., "Impact of Land Use Changes on Transportation Networks", Canadian Journal of Civil Engineering, 3 (3): 372-278, September 1975.

Sosslau, A.B., et al., Travel Estimation Procedures for Quick to Urban Policy Issues (Draft), Volume II, prepared for the Transportation Research Board, Washington, DC, February 1977.

Tardiff, T.J., "Trip Generation as a Choice Process", ASCE Transportation Engineering Journal, 103 (2): 337-348, March 1977.

Transportation Research Board, Research Record #569, Washington, DC, 1976, NTIS PB-256 975/4ST.

U.S. Environmental Protection Agency, Land Use Information for Water Quality Management Planning, Washington, DC, August 1976.

Weiner, E., Kassoff, H., and Gendell, D.S., "Multimodel National Urban Transportation Policy Planning Model", Highway Research Record #458, Highway Research Board, Washington, DC, 1973.

Zaryouni, M.R. and Kannel, E.J., "Synthesized Trip-Forecasting Model for Small and Medium-Sized Urban Areas".

## 3. Land Use Identification and Categorization from Aerial Photographs

Adams, V.W., 1975, Earth Science Data in Urban and Regional Information Systems--A Review, US Geol. Survey Circ. 712.

Aguilar, A.M., "Area & Volume Errors in Reservoir Projects", ASCE Journal of Surveying & Mapping, pp. 287-296, November 1971.

American Society of Photogrammetry, Manual of Photographic Interpretation, Washington, DC, American Society of Photogrammetry, 1960.

Anderson, J.R., Hardy, E.E., and Roach, J.T., 1972, *A Land-Use Classification System for Use with Remote-Sensor Data*, U.S. Geol. Survey Circ. 671.

Anderson, J.R., Hardy, E.E., Roach, J.T., and Witmer, R.E., *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*, USGS Paper #964, Washington, DC, 1976.

Aschenbrenner, C.M., "Problems in Getting Information into and out of Air Photographs", *Photogrammetric Engineering*, *20* (3): 398-401, 1954.

Avery, T. Eugene, *Interpretation of Aerial Photographs*, Minneapolis, MN, Burgess Publishing Co., 1968.

Branch, Melville C., *City Planning and Aerial Information*, Cambridge, MA, Harvard University Press, 1971.

Branch, Melville C., *Planning Urban Environment*, Dowden, Hutchinson & Roos, Inc., 1974, Stroudsburg, PA, 1974.

Caspan Corp., *Remote Sensing for land use analysis*, Houston, TX, June 1975, NTIS N75-31938.

Collins, W.G., and El-Beik, A.H., "The Acquisition of Urban Land Use Information from Aerial Photographs of the City of Leeds (Great Britain)", *Photogrammetria*, *27*:71, 1971.

Colwell, R.N., "A Systematic Analysis of Some Factors Affecting Photographic Interpretations", *Photogrammetric Engineering*, *20* (3):433-454, 1954.

Eardley, A.J., *Aerial Photographs--Their Use and Interpretation*, Harper and Brothers, Publishers, New York, NY, 1942.

Environmental Research Institute of Michigan, *Proceedings of the Tenth International Symposium on Remote Sensing of Environment*, Ann Arbor, MI, October 6-10, 1975.

Fitzpatrick, K.A., *Cost, Accuracy and Consistency Comparisons of Land Use Maps Made from High-Altitude Aircraft Photography and ERTS Imagery*, U.S. Geological Survey, Reston, VA, September 1975.

Gautum, N.C., "Aerial Photo-Interpretation Techniques for Classifying Urban Land Use", *Photogrammetric Engineering and Remote Sensing*, *42*(6):815-822, 1976.

Hardy, Ernest E., Belcher, Donald J., and Phillips, Elmer S., *Land Use Classification with Simulated Satellite Photography*, U.S. Department of Agriculture Economic Research Service, Agricultural Information Bulletin, Washington, DC, 1971.

Howard, William A., and Kraht, James B., An Assessment of the Usefulness of Small-Scale Photographic Imagery for Acquiring Land Use Information Necessary to the Urban Planning Function, Technical Paper No. 71-2, Denver, CO, Department of Geography, University of Denver, September 1971, mimeo.

Lafferty, Maurice E., "Accuracy/Costs with Analysis", Photogrammetric Engineering, 39 (5): 507-510, May 1975.

Lehman, E.J., Remote Sensing for natural resource, environmental and regional planning, January 1975, NTIS PS-75/104.

Lindren, D.T., Dwelling Unit Estimation from Color Infrared Photography, Dartmouth College, Department of Geography, Paper No. 1, Hanover, NH, May 1970.

Lins, H.F., "Land-Use Mapping from Skylab S-190B Photography", Photogrammetric Engineering and Remote Sensing, 42(3):301-307, March 1976.

Lueder, Donald R., Aerial Photographic Interpretation: Principles and Applications, McGraw-Hill, New York, NY, 1959.

Mausel, P.W., Leiro, C.E., and Lewellen, M.T., "Regional Land Use Classification Derived from Computer-Processed Satellite Data", Journal of the American Institute of Planners, 42:153-164, April 1974.

Muret, Jean-Pierre, Aerial Photography and City Planning, Paris, France, Centre de Recherche d'Urbanisme, International Federation for Housing and Planning, 1972.

Perlman, E. and Raney, R., An Experiment in the Application of Remote Sensing to Land-Use Planning on the Urban Fringe, Environmental Research Institute of Michigan, Ann Arbor, MI, June 1975.

Ray, R.G., Aerial Photographs in Geologic Interpretation and Mapping, Geological Survey Paper #373, Washington, DC, 1960.

Simpson, R.B. and Lindgren, D.T., Recognition of Settlement Patterns Against a Complex Background, U.S. Geological Survey, Washington, DC, NTIS N71-3430, May 1970.

Smith, H.T.U., Aerial Photographs and Their Applications, New York, NY. Appleton-Century-Crofts, Inc., 1943.

Soil Conservation Service, Aerial-Photo Interpretation in Classifying and Mapping Soils, U.S. Dept. of Agriculture Handbook #294, Washington, DC, October 1966.

Spangle, W. and Associates; Leighton, F.B. and Associates; and Baxter, McDonald and Company, 1976, Earth Science Information in Land-Use Planning, U.S. Geol. Survey Circ. 721.

Stafford, D.A., "Measuring Watershed Land-Use Changes with Air-photos", ASCE Transportation Enginerring Journal, pp. 117-129, February 1976.

Stone, Kirk, "Air Photo Interpretation Procedures", Photogrammetric Engineering, 22(1):123-132, 1951.

Summerson, C.H., "A Philosophy for Photo Interpreters", Photogrammetric Engineering, 20(3):396-397, 1954.

Urban Renewal Administration, Standard Land Use Coding Manual, Washington, DC, January 1965.

U.S. Dept. of Interior, Proceedings of the National Conference on Land Use Information and Classification, June 28-30, 1971,

Weedel, J.W. and Klickner, R., Using remote data for land use mapping and inventory: a user guide, Association of American Geographics, Washington, DC, July 1974, NTIS PB-242-813.

Wellar, Barry S., Hyperaltitude Photography as a Data Base in Urban and Regional Planning, Evanston, Ill, Remote Sensing Laboratory, Department of Geography, Northwestern University, January 1969.

### 4. Potential Case Study Wastewater Projects

Abrams, P., "Plant Load Jumps Fourfold as Recreationists Come to Town", Wastes Engineering, 32(11):672-675, November 1961.

Barksdale, W.D., "Modern is Word for Weirton Sewage Plant", American City, 76(2):100-103, February 1961.

Bascome, S.E., Cooper, K.G., Howell, M.P., Makrides, A.C., and Rabe, F.T., Secondary Impacts of Transportaion and Wastewater Investments: Research Results (EPA-600/5-75-013) and Review and Bibliography (EPA-600/5-75-002), Washington, DC, 1975.

Bi-State Metropolitan Planning Commission, Metropolitan Comprehensive Water Sewage, and Solid Waste Planning Study (3 Volumes), Rock Island, IL, March 1970.

Bower, B., et al., Waste Management - Generation and Disposal of Solid, Liquid, and Gaseous Wastes in the New York Region (A Report of the Second Regional Plan), Regional Plan Association, Inc., New York, NY, March 1968.

Browner, H.D., "Semi-elliptical Sewer", Civil Engineering, 29(6): 77-79, June 1959.

Bureau of Regional Planning, Secondary Impact of Regional Sewerage Systems, Trenton, NJ, State of New Jersey Department of Community Affairs, 1975.

"Camouflaged Sewage Treatment", American City, 74(5):104-1066, May 1959.

Camp, Dresser, & McKee, Inventory of Water and Sewer Facilities - Eastern Massachusetts Regional Planning Project, Boston, MA, May 1967.

Cawley, W.A., "Mill Creek Sewage Works", Water & Sewage Works, 106(10):436-439, October 1959.

Central Iowa Regional Planning Commission, Des Moines Metropolitan Sanitary Sewerage System Study, December 1971.

Cobb, E.B. and Wheeler, H.R., "Progress Report on Development of Allegheny County Sanitary Authority Facilities of Pittsburgh, Pa.", Journal of Boston Society of Civil Engineers, 46(1):38-56, January 1959.

Denver Regional Council of Governments, Drainage Basin Descriptions Sub-basin Delineation and Summaries, Denver, CO, March 1972, NTIS No. PB-219-610.

East-West Gateway Coordinating Council, St. Louis Region Water and Sewer Report - Five-Year Program, St. Louis, MO, June 1973, NTIS No. PB-222-723.

Graeser, H.J., "Expanded Sewage Treatment for Dallas", Water & Sewage Works, 106(10):411-415, October 1959.

Hahn, H.H., "Regional Wastewater Management Systems", Modes for Environmental Pollution Control, Ann Arbor Science Publishers, pp. 41-60, 1973.

Hayes, G.G., Institutional Alternatives for Providing Programmed Water and Sewer Services in Urban Growth Areas: A Case Study of Knoxville - Knox County, Tenn., Knoxville Water Resources Research Center, Tennessee University Report No. 18, June 1972.

Hensley - Schmidt, Inc., Chattanooga Area Regional Council of Governments Comprehensive Water and Sewer Study: Hamilton, Walker and Catoosa Counties, Chattanooga, TN, December 1971.

Houston-Galveston Regional Transportation Study, Population and Land Use 1970-1990, Houston, TX, July 1974.

Howells, D.H., Proceedings of Symposium on Better Water and Sewer Services for Small Communities in North Carolina, Water Resources Research Institute, North Carolina University, December 1968.

Karfman, H.L., "Mechanization Featured at New Primary Plant for Binghamton", Wastes Engineering, 32(4):190-192, April 1961.

Kentucky Office for Local Government, Modified and Updated Comprehensive Water and Sewer Plan for Green River and Purchase Area Development Districts, Frankfort, KY, 1973, NTIS Nos. PB-223-058 and PB-223-045.

Laboon, J.F., "Construction and Operation of Pittsburgh Project", Journal of Water Pollution Control Federation, 33(7):758-782, July 1961.

Lexington-Fayette County Planning Commission, A Growing Community, Lexington, KY, 1973.

Loucks, D.P., "Annual Literature Review: Administration, Systems Analysis", Journal Water Pollution Control Federation, 46: 1604-1611, June 1974.

McWilliams, W., "Denver's Sewage Disposal Problem", Water & Sewage Works, 107(7):287-290, July 1960.

Metro Dade County Planning Dept., Land Use Characteristics: 1960, 1970, Miami, Florida, July 1973.

Metropolitan Washington Council of Governments, Water and Sewer Plan Program - 1970, Washington, DC, September 1970.

Middle Flint Area Planning and Development Commission, Middle Flint Regional Water and Sewer Systems Plan, Ellaville, GA, March 1973, NTIS No. PB-221-528.

Mid-Missouri Regional Planning Commission, Phase I Comprehensive Water and Sewer Plan, Jefferson City, MO, September 1971.

Minnesota Pollution Control Agency, Wastewater Disposal Facility Inventory, State of Minnesota, Division of Water Quality, Roseville, MN, July 1975.

Murray, C.E., "One Way to Get Sewage Plant Built: Prohibit Further Sewer Extensions", Wastes Engineering, 31(3):138-140, March 1960.

Oklahoma Foundation for Residential & Development Utilization, Inc., Preliminary Listing of Municipal Wastewater Treatment Capacities, Economic Development Administration, U.S. Department of Commerce, September 1976 (NTIS #PB-254-430).

San Diego County Comprehensive Planning Organization, Water Distribution and Sanitary Sewerage Systems Background and Policy Study, San Diego, CA, February 1972.

Scott, G.R., "New Sewage Treatment Works for Boulder, CO", American City, 74(1):129-131, January 1959.

Southwestern Illinois Metropolitan Area Planning Commission, *Comprehensive Water and Sewer Plan - Randolph County*, Collinville, IL, December 1972, NTIS No. PB-216-396.

Top of Alabama Regional Council of Governments, *Regional Land Use Survey and Analyses*, Huntsville, AL, May 1973.

Townsend, J.W. and Beckman, J.J., "Pleasant Hills Sewage Plant", *Water & Sewage Works*, *107*(10):370-376, October 1960.

Tri-State Regional Planning Commission, *The Extent of Public Water and Sewer Systems in the Tri-State Region*, New York, NY, May 1973.

U.S. Bureau of the Census, *Census of Population and Housing 1970*, Fourth-Count Housing Tallies.

U.S. Dept. of Commerce, *Accelerated Public Works Program Directory of Approved Projects as of July 1, 1964*, Area Redevelopment Administration, Washington, DC, August 1964.

U.S. Environmental Protection Agency, *1968 Inventory, Municipal Waste Facilities*, Volumes I-X for each federal region, Office of Water Programs, Washington, DC, 1971 (EPA Publication No. OWP-1).

U.S. Environmental Protection Agency, *Project Register - Wastewater Treatment Construction Grants*, Grants Administration Division, Washington, DC, June 30, 1973.

U.S. Environmental Protection Agency, Region II, *Wastewater Treatment Facilities Grants for Nassau and Suffolk Counties*, July 1972.

U.S. Environmental Protection Agency, Region III, *Bethany Beach (Delaware) Regional Wastewater Treatment Plant*, December 1972.

U.S. Environmental Protection Agency, Region IV, *Cobb County (Georgia) Sewerage Improvement Project*, July 1971.

U.S. Environmental Protection Agency, Region VI, *Wastewater Facilities, Hot Springs, Arkansas*, October 1972.

U.S. Environmental Protection Agency, Region VIII, *Denver Regional Environmental Statement for Wastewater Facilities and the Clean Water Plan* (Draft), Denver, CO. June 1977.

U.S. Environmental Protection Agency, *Wastewater Contract Awards*, Computer data file maintained by the Municipal Construction Division.

U.S. Public Health Service, "Sewer and Water Works Construction, 1960", publication number 758, 1961.

Southwestern Illinois Metropolitan Area Planning Commission, *Comprehensive Water and Sewer Plan - Randolph County*, Collinville, IL, December 1972, NTIS No. PB-216-396.

Top of Alabama Regional Council of Governments, *Regional Land Use Survey and Analyses*, Huntsville, AL, May 1973.